

Assignment05

March 6, 2022

1 Assignment 5

1.1 Web scraping

Use requests and BeautifulSoup to make a list of all the CORGIS datasets. - Use requests with the following site: “<https://corgis-edu.github.io/corgis/csv/>” - The relevant HTML tag to search for is “h3”

```
[1]: import requests
```

```
[2]: response = requests.get('https://corgis-edu.github.io/corgis/csv/')
response
```

```
[2]: <Response [200]>
```

```
[3]: response.text
```

```
[3]: '<!DOCTYPE html>\n<html lang="en-US">\n    <head>\n        <meta charset="utf-8">\n        <meta http-equiv="X-UA-Compatible" content="IE=edge">\n        <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">\n        \n        \n        <title>CORGIS Datasets Project</title>\n        <meta name="description" content="\n        \n        \n        \n        \n        Github Pages for CORGIS Datasets Project">\n        <link rel="canonical" href="https://corgis-edu.github.io//corgis/csv/">\n        <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css">\n        <link rel="stylesheet" href="https://code.jquery.com/ui/1.12.1/themes/base/jquery-ui.css">\n        <link rel="stylesheet" href="/corgis/assets/css/main.css">\n        \n        \n        \n        <script src="/corgis/assets/js/jquery-3.3.1.min.js"></script>\n        <script src="https://code.jquery.com/ui/1.12.1/jquery-ui.min.js"\n        integrity="sha256-VazP97ZCwtekAsvgPBSUwPFKdrwD3unUfSGVYrahUqU="\n        crossorigin="anonymous"></script>\n        \n        <script src="https://cdnjs.cloudflare.com/ajax/libs/knockout/3.5.0/knockout-debug.js"\n        integrity="sha256-M5ZomNNrnEB2WjSbnty5GWGqq6UuAAVnNwECisgEis="\n        crossorigin="anonymous"></script>\n        \n        <link rel="stylesheet" href="/corgis/assets/css/multiselect.css">\n        <link rel="stylesheet" href="/corgis/assets/css/bootstrap-tagsinput.css">\n        \n        <script src="/corgis/assets/js/knockout-3.5.0.js"></script>\n        <script src="/corgis/assets/js/d3.min.js"></script>\n        <script
```

```

src="/corgis/assets/js/jquery.multi-select.js"></script>\n    <script
src="/corgis/assets/js/bootstrap-tagsinput.js"></script>\n    <script
src="/corgis/assets/js/typeahead.bundle.js"></script>\n    <script
src="/corgis/assets/js/jquery.quicksearch.js"></script>\n    <script
src="/corgis/assets/js/jquery.multi-select.js"></script>\n    <script
src="/corgis/assets/js/lunr.js"></script>\n\n    \n\n</head>\n    <body
class="layout--default "\n    >\n    <div class="container-fluid">\n    \n\n\n<script>\nlet searcher = lunr(function() {\n    this.field('\name', 10);\n    this.field('\tags', 100);\n    this.field('\body');\n    this.field('\description'); \n    this.ref('\name');\n});\n</script>\n\n<p><a
href="/corgis/" class="site-logo" rel="home" title="CORGIS Datasets Project">\n
\n</a></p>\n\n<h1 id="csv-datasets">CSV
Datasets</h1>\n\n<p class="lead">CORGIS: The Collection of Really Great,
Interesting, Situated Datasets</p>\n\n<p><span class="text-muted">By Austin Cory
Bart, Ryan Whitcomb, Jason Riddle, Omar Saleem, Dr. Eli Tilevich, Dr. Clifford
A. Shaffer, Dr. Dennis Kafura</span></p>\n\n<hr />\n\n<form>\n<div class="form-
group row search-bar">\n    <label for="search-bar" class="col-sm-1 col-form-
label">Search:</label>\n    <div class="col-sm-11">\n        <input type="text"
class="form-control" id="search-bar" placeholder="Keyword or phrase" />\n
</div>\n</div>\n</form>\n\n<hr />\n\n<div class="row">\n\n\n\n<div class="col-4
ml-md-auto dataset-card" id="aids">\n\n    <a href="aids/" class="dataset-link
float-left">\n\n        \n\n        </a>\n\n        <a href="aids/" class="dataset-
link">\n            <h3>Aids</h3>\n            </a>\n\n            <p>\n                Records of AIDS related
statistics from several countries.\n                <a href="aids/" class="dataset-
link"><em>View</em></a>\n                </p>\n\n                <small>aids, death, disease, hiv,
orphans, health, countries,...</small>\n\n                <script>\n                    searcher.add(\n
{\n                        \name': "aids",\n                        \body': "Records of AIDS related
statistics from several countries.",\n                        \description': "The UNAIDS
Organization is an entity of the United Nations that looks to reduce the
transmission of AIDS and provide resources to those currently affected by the
disease. The following data set contains information on the number of those
affected by the disease, new cases of the disease being reported, and AIDS-
related deaths for a large set of countries over the course of 1990 - 2015.",\n
\tags': ["aids", " death", " disease", " hiv", " orphans", " health", "
countries", " world", " gender", " united nations", " un"]\n                    }\n                );\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="airlines">\n\n    <a href="airlines/" class="dataset-link float-left">\n\n\n
\n\n        </a>\n\n        <a href="airlines/" class="dataset-link">\n
<h3>Airlines</h3>\n        </a>\n\n        <p>\n            Information about flight delays
in major airports since 2003.\n            <a href="airlines/" class="dataset-
link"><em>View</em></a>\n            </p>\n\n            <small>airplane, airports, travel,
plane, air, flights, delays,...</small>\n\n            <script>\n                searcher.add(\n
{\n                    \name': "airlines",\n                    \body': "Information about
flight delays in major airports since 2003.",\n                    \description': "This

```

dataset is all about flights in the united states, including information about the number, length, and type of delays. The data is reported for individual months at every major airport for every carrier. Additional information is available: <http://www.rita.dot.gov/bts/help/aviation/html/understanding.html>",\n\n\tags\': ["airplane", " airports", " travel", " plane", " air", " flights", " delays", " national", " united states", " transportation"]\n\n\t);\n\n</script>\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card" id="billionaires">\n\n\t\n\n\t\t\n\n\t\t\n\n\t\t\t<h3>Billionaires</h3>\n\n\t\t\n\n\t<p>\n\n\t\tInformation about over 2000 billionaires from around the world.\n\n\t\tView\n\n\t\t</p>\n\n\t<small>money, rich, wealthy, people, person, billionaire</small>\n\n\t<script>\n\n\t\tsearcher.add(\n\n\t\t\t{\n\n\t\t\t\t'name\': "billionaires",\n\n\t\t\t\t'body\': "Information about over 2000 billionaires from around the world.",\n\n\t\t\t\t'description\': "Researchers have compiled a multi-decade database of the super-rich. Building off the Forbes World's Billionaires lists from 1996-2014, scholars at Peterson Institute for International Economics have added a couple dozen more variables about each billionaire - including whether they were self-made or inherited their wealth. (Roughly half of European billionaires and one-third of U.S. billionaires got a significant financial boost from family, the authors estimate.)",\n\n\t\t\t\t'tags\': ["money", " rich", " wealthy", " people", " person", " billionaire"]\n\n\t\t\t}\n\n\t\t);\n\n\t</script>\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card" id="broadway">\n\n\t\n\n\t\t\n\n\t\t\n\n\t\t\t<h3>Broadway</h3>\n\n\t\t\n\n\t<p>\n\n\t\tThis library holds data about Broadway shows, such as tickets sold.\n\n\t\tView\n\n\t</p>\n\n\t<small>broadway, musical, theatre, tickets</small>\n\n\t<script>\n\n\t\tsearcher.add(\n\n\t\t\t{\n\n\t\t\t\t'name\': "broadway",\n\n\t\t\t\t'body\': "This library holds data about Broadway shows, such as tickets sold.",\n\n\t\t\t\t'description\': "This library holds data about over Broadway shows, grouped over weeklong periods. Only shows that reported capacity were included, so the dataset stretches back to the 1990s. The dataset is made available by the Broadway League (the national trade association for the Broadway industry), and you can view the data online at <http://www.broadwayleague.com/>.",\n\n\t\t\t\t'tags\': ["broadway", " musical", " theatre", " tickets"]\n\n\t\t\t}\n\n\t\t);\n\n\t</script>\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card" id="business dynamics">\n\n\t\n\n\t\t\n\n\t\t\n\n\t\t\t<h3>Business Dynamics</h3>\n\n\t\t\n\n\t<p>\n\n\t\tThe Business Dynamics Statistics

(BDS) includes measures of establishment openings and closings, firm startups, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics for the US.

[View](business_dynamics/)

government, united states, us, usa, business, businesses,...

```

<script>\n    searcher.add(\n        {\n            \'name\':
"business_dynamics",\n            \'body\': "The Business Dynamics Statistics
(BDS) includes measures of establishment openings and closings, firm startups,
job creation and destruction by firm size, age, and industrial sector, and
several other statistics on business dynamics for the US.",\n            \'description\': "The Business Dynamics Statistics (BDS) includes measures of
establishment openings and closings, firm startups, job creation and destruction
by firm size, age, and industrial sector, and several other statistics on
business dynamics. The U.S. economy is comprised of over 6 million
establishments with paid employees. The population of these businesses is
constantly churning -- some businesses grow, others decline and yet others
close. New businesses are constantly replenishing this pool. The BDS series
provide annual statistics on gross job gains and losses for the entire economy
and by industrial sector, state, and MSA. These data track changes in employment
at the establishment level, and thus provide a picture of the dynamics
underlying aggregate net employment growth. There is a longstanding interest in
the contribution of small businesses to job and productivity growth in the U.S.
Some recent research suggests that it is business age rather than size that is
the critical factor. The BDS permits exploring the respective contributions of
both firm age and size. BDS is based on data going back through 1976. This
allows business dynamics to be tracked, measured and analyzed for young firms in
their first critical years as well as for more mature firms including those that
are in the process of reinventing themselves in an ever changing economic
environment. If you need help understanding the terms used, check out these
definitions.",\n            \'tags\': ["government", " united states", " us", "
usa", " business", " businesses", " firms", " establishments", " jobs", "
census"]\n        }\n    );\n    </script>\n\n\n\n</div>\n\n\n<div class="col-4
ml-md-auto dataset-card" id="cancer">\n\n    <a href="cancer/" class="dataset-
link float-left">\n\n        \n\n        </a>\n\n        <a href="cancer/"
class="dataset-link">\n            <h3>Cancer</h3>\n            </a>\n\n        <p>\n
Cumulative cancer deaths for the period 2007-2013 are reported for each U.S.
state.\n            <a href="cancer/" class="dataset-link"><em>View</em></a>\n
</p>\n\n        <small>cancer, cancer deaths, medical, health</small>\n\n
<script>\n    searcher.add(\n        {\n            \'name\': "cancer",\n
\'body\': "Cumulative cancer deaths for the period 2007-2013 are reported for
each U.S. state.",\n            \'description\': "Information about the rates of
cancer deaths in each state is reported. The data shows the total rate as well
as rates based on sex, age, and race. Rates are also shown for three specific
kinds of cancer: breast cancer, colorectal cancer, and lung cancer.",\n            \'tags\': ["cancer", " cancer deaths", " medical", " health"]\n        }\n    );\n    </script>\n\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"

```

```

id="cars">\n\n    <a href="cars/" class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a href="cars/" class="dataset-link">\n
<h3>Cars</h3>\n    </a>\n\n    <p>\n        This is a dataset about cars and how
much fuel they use.\n        <a href="cars/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>cars, vehicles, fuel</small>\n\n
<script>\n    searcher.add(\n        {\n            \'name\': "cars",\n
\'body\': "This is a dataset about cars and how much fuel they use.",\n
\'description\': "This is a dataset about cars and how much fuel they use.",\n
\'tags\': ["cars", " vehicles", " fuel"]\n        }\n    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="classics">\n\n    <a href="classics/" class="dataset-link float-left">\n\n
\n\n    </a>\n\n    <a href="classics/" class="dataset-link">\n
<h3>Classics</h3>\n    </a>\n\n    <p>\n        Records and computed statistics
about the top 1000 books on Project Gutenberg.\n        <a href="classics/"
class="dataset-link"><em>View</em></a>\n    </p>\n\n    <small>classics, books,
texts, text, book, classic, english,...</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "classics",\n
\'body\': "Records and computed statistics about the top 1000 books on Project
Gutenberg.",\n            \'description\': "Project Gutenberg (PG) is a
volunteer effort to digitize and archive cultural works, to \'encourage the
creation and distribution of eBooks\'. It was founded in 1971 by Michael S. Hart
and is the oldest digital library. This dataset is a collection of the top 1000
most popular books on Project Gutenberg, as determined by downloads. Each book
has information about its authorship, publication date, congressional
classification, and a few other fields. It also has some simple, computed
statistics based on common metrics such as sentiment analysis, Flesch Kincaid
Reading level, and average sentence length.",\n            \'tags\':
["classics", " books", " texts", " text", " book", " classic", " english", "
shakespeare", " literature", " novel", " language", " composition", " writing",
" author", " publication", " words"]\n        }\n    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="construction permits">\n\n    <a href="construction_permits/"
class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a
href="construction_permits/" class="dataset-link">\n    <h3>Construction
Permits</h3>\n    </a>\n\n    <p>\n        This dataset provides data on the
number and valuation of new housing units authorized by building permits.\n
<a href="construction_permits/" class="dataset-link"><em>View</em></a>\n
</p>\n\n    <small>construction, permits, us, united states,
america,...</small>\n\n    <script>\n    searcher.add(\n        {\n
\'name\': "construction_permits",\n            \'body\': "This dataset provides
data on the number and valuation of new housing units authorized by building
permits.",\n            \'description\': "This dataset provides data on the
number and valuation of new housing units authorized by building permits. Data

```

are available monthly at the national, regional, divisional, and state level. The data are from the Building Permits Survey. The United States Code, Title 13, authorizes this survey, provides for voluntary responses, and provides an exception to confidentiality for public records.",\n

```

\tags\':
["construction", " permits", " us", " united states", " america", " government",
" buildings", " private", " units", " valuation", " housing"]\n
}\n
);\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="construction spending">\n\n
<a href="construction_spending/"
class="dataset-link float-left">\n\n
\n\n
</a>\n\n
<a
href="construction_spending/" class="dataset-link">\n
<h3>Construction
Spending</h3>\n
</a>\n\n
<p>\n
Estimates of the total dollar value
of construction work done in the U.S.\n
<a href="construction_spending/"
class="dataset-link"><em>View</em></a>\n
</p>\n\n
<small>construction,
spending, us, united states, america,...</small>\n\n
<script>\n
searcher.add(\n
{\n
\t'name\': "construction_spending",\n
\t'body\': "Estimates of the total dollar value of construction work done in the
U.S.",\n
\t'description\': "The Value of Construction Put in Place
Survey (VIP) provides monthly estimates of the total dollar value of
construction work done in the U.S. The United States Code, Title 13, authorizes
this program. The survey covers construction work done each month on new
structures or improvements to existing structures for private and public
sectors. Data estimates include the cost of labor and materials, cost of
architectural and engineering work, overhead costs, interest and taxes paid
during construction, and contractor\'s profits. Data collection and estimation
activities begin on the first day after the reference month and continue for
about three weeks. Reported data and estimates are for activity taking place
during the previous calendar month. The survey has been conducted monthly since
1964. Why isn\'t the \'total construction\' the sum of the other categories? It
is! the trick is that the \'residential\' and \\'nonresidential\' categories
are supposed to be below the \'total construction\' level, and the other
categories are below those two categories. This example of the data should make
it more clear: http://www.census.gov/construction/c30/pdf/tot.pdf.More
information about the nature of the construction projects is also available.",\n
\tags\': ["construction", " spending", " us", " united states", " america", "
government", " buildings", " residential", " non-residential", " private", "
public"]\n
}\n
);\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4
ml-md-auto dataset-card" id="county demographics">\n\n
<a
href="county_demographics/" class="dataset-link float-left">\n\n
\n\n
</a>\n\n
<a
href="county_demographics/" class="dataset-link">\n
<h3>County
Demographics</h3>\n
</a>\n\n
<p>\n
Demographic information for
counties in the United States.\n
<a href="county_demographics/"
class="dataset-link"><em>View</em></a>\n
</p>\n\n
<small>demographics,
population, age, ethnicity, race, language,...</small>\n\n
<script>\n

```

```

searcher.add(\n          {\n          \'name\': "county_demographics",\n          \'body\': "Demographic information for counties in the United States.",\n          \'description\': "The following data set is information obtained about counties in the United States from 2010 through 2019 through the United States Census Bureau. Information described in the data includes the age distributions, the education levels, employment statistics, ethnicity percents, household information, income, and other miscellaneous statistics. (Values are denoted as -1, if the data is not available)",\n          \'tags\': ["demographics", "population", "age", "ethnicity", "race", "language", "education", "foreign", "households", "income", "employment", "unemployment", "sales", "land area", "counties"]\n          });\n</script>\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card" id="covid">\n  <a href="covid/" class="dataset-link float-left">\n  \n  </a>\n  <a href="covid/" class="dataset-link">\n    <h3>Covid</h3>\n  </a>\n  <p>\n    Since the beginning of the coronavirus pandemic, the Epidemic Intelligence team of the European Center for Disease Control and Prevention (ECDC) has been collecting on daily basis the number of COVID-19 cases and deaths, based on reports from health authorities worldwide. This helps to monitor and interpret the dynamics of the COVID-19 pandemic not only in the European Union (EU), the European Economic Area (EEA), but also worldwide. Every day a team of epidemiologists screens up to 500 relevant sources to collect the latest figures. The data screening is followed by ECDC standard epidemic intelligence process for which every single data entry is validated and documented in an ECDC database.\n    <a href="covid/" class="dataset-link"><em>View</em></a>\n  </p>\n  <small>Covid, Covid-19, pandemic, infection, world health</small>\n  <script>\n    searcher.add(\n    {\n      \'name\': "covid",\n      \'body\': "Since the beginning of the coronavirus pandemic, the Epidemic Intelligence team of the European Center for Disease Control and Prevention (ECDC) has been collecting on daily basis the number of COVID-19 cases and deaths, based on reports from health authorities worldwide. This helps to monitor and interpret the dynamics of the COVID-19 pandemic not only in the European Union (EU), the European Economic Area (EEA), but also worldwide. Every day a team of epidemiologists screens up to 500 relevant sources to collect the latest figures. The data screening is followed by ECDC standard epidemic intelligence process for which every single data entry is validated and documented in an ECDC database.",\n      \'description\': "The data set contains daily reports of Covid-19 cases and deaths in countries worldwide. The data also shows the country\'s population and the number of cases per 100,000 people on a rolling 14 day average.",\n      \'tags\': ["Covid", "Covid-19", "pandemic", "infection", "world health"]\n    });\n  </script>\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card" id="drugs">\n  <a href="drugs/" class="dataset-link float-left">\n  \n  </a>\n  <a href="drugs/" class="dataset-link">\n    <h3>Drugs</h3>\n  </a>\n  <p>\n    This dataset is about substance abuse (cigarettes, marijuana, cocaine, alcohol) among different age groups and

```

```

states.\n      <a href="drugs/" class="dataset-link"><em>View</em></a>\n
</p>\n\n      <small>drugs, substances, drug usage</small>\n\n      <script>\n
searcher.add(\n          {\n              \'name\': "drugs",\n              \'body\':
"This dataset is about substance abuse (cigarettes, marijuana, cocaine, alcohol)
among different age groups and states.",\n              \'description\': "This
dataset is about substance abuse (cigarettes, marijuana, cocaine, alcohol) among
different age groups and states. Data was collected from individual states as
part of the NSDUH study. The data ranges from 2002 to 2018. Both totals (in
thousands of people) and rates (as a percentage of the population) are
given.",\n              \'tags\': ["drugs", " substances", " drug usage"]\n
          });\n      </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto
dataset-card" id="earthquakes">\n\n      <a href="earthquakes/" class="dataset-
link float-left">\n\n          \n\n          <a
href="earthquakes/" class="dataset-link">\n              <h3>Earthquakes</h3>\n
</a>\n\n          <p>\n              Records from different earthquake occurences across the
world.\n              <a href="earthquakes/" class="dataset-link"><em>View</em></a>\n
</p>\n\n          <small>earthquakes, nature, disaster, magnitude, richter
scale,...</small>\n\n          <script>\n              searcher.add(\n                  {\n
\'name\': "earthquakes",\n                  \'body\': "Records from different
earthquake occurences across the world.",\n                  \'description\':
"Earthquake records from around the world collected from the United States
Geological Survey. Important details about the earthquake such as distance, gap,
magnitude, depth and significance are included to properly describe the
earthquake. Additionally, data about exact geological coordinates and a relative
description of the earthquake\'s location is included. The earthquakes collected
are from the past month.",\n                  \'tags\': ["earthquakes", " nature", "
disaster", " magnitude", " richter scale", " latitude", " longitude"]\n
              });\n          </script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto
dataset-card" id="election">\n\n          <a href="election/" class="dataset-link
float-left">\n\n              \n\n              <a
href="election/" class="dataset-link">\n                  <h3>Election</h3>\n
</a>\n\n              <p>\n                  A breakdown of how each county voted in the 2016 Presidential
primaries.\n                  <a href="election/" class="dataset-link"><em>View</em></a>\n
</p>\n\n              <small>politics, election, votes, counties, candidate</small>\n\n
<script>\n                  searcher.add(\n                      {\n
\'name\': "election",\n                      \'body\': "A breakdown of how each county voted in the 2016 Presidential
primaries.",\n                      \'description\': "Data about the results was collected
from CNN\'s 2016 election center. Each county has the breakdown of how many
votes each candidate recieved in that county and what percentage of the total
vote from that county that the candidate recieved. The candidate\'s political
party is also represented.",\n                      \'tags\': ["politics", " election", "
votes", " counties", " candidate"]\n                      });\n                  </script>\n\n\n\n\n</div>\n\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="electricity">\n\n          <a href="electricity/" class="dataset-link float-
left">\n\n              <img src="/corgis//images/datasets/power-line-icon.png"

```


alt="Electricity" class="img-thumbnail m-2" />\n\n \n\n \n <h3>Electricity</h3>\n \n\n <p>\n This data set describes over 2000 U.S. electric utilities. Reported data for 2019 includes electrical generation, distribution, revenues, and customers.\n View\n </p>\n\n <small>elctricity,electric power, utilities, utility revenues</small>\n\n <script>\n searcher.add(\n {\n \'name\': "electricity",\n \'body\': "This data set describes over 2000 U.S. electric utilities. Reported data for 2019 includes electrical generation, distribution, revenues, and customers.",\n \'description\': "The U.S. Energy Information Adminstration collects and curates self-reported information from energy utilities about energy production and usage in the United States. This data set contains information from over 2,000 U.S. utilities in 2017. The information includes sources of energy, its uses in different economic sectors, and the revenues obtained from the sale of electrical energy.",\n \'tags\': ["elctricity", "electric power", "utilities", " utility revenues"]\n });\n </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card" id="emissions">\n\n \n\n \n\n \n\n \n <h3>Emissions</h3>\n \n\n <p>\n The Emissions Database for Atmospheric Research (EDGAR) supported by the European Union shows green house gas emissons by country. This data set contains data from 1970 through 2012. \n View\n </p>\n\n <small>hospitals, health care, medical, hospital costs, hospital quality</small>\n\n <script>\n searcher.add(\n {\n \'name\': "emissions",\n \'body\': "The Emissions Database for Atmospheric Research (EDGAR) supported by the European Union shows green house gas emissons by country. This data set contains data from 1970 through 2012.",\n \'description\': "The data set allows consumers to directly compare across hospitals performance measure information related to heart attack, emergency department care, preventive care, stroke care, and other conditions. The data is part of an Administration-wide effort to increase the availability and accessibility of information on quality, utilization, and costs for effective, informed decision-making.",\n \'tags\': ["hospitals", " health care", " medical", " hospital costs", " hospital quality"]\n });\n </script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card" id="energy">\n\n \n\n \n\n \n\n \n <h3>Energy</h3>\n \n\n <p>\n United States Government reports on consumption, production, import, and export of various fuel sources.\n View\n </p>\n\n <small>energy, renewable, united states, us, nuclear, fossil fuels, power</small>\n\n <script>\n searcher.add(\n {\n

```

\'name\': "energy",\n          \'body\': "United States Government reports on
consumption, production, import, and export of various fuel sources.",\n
\'description\': "United States Government reports on consumption, production,
expenditures, and prices of various fuel sources. This data comes from the US
Energy Information Administration, which has historical data from 1960 to 2019.
Information was not always available, in which case 0 was reported instead. In
some cases, very tiny values were rounded down to zero.",\n          \'tags\':
["energy", " renewable", " united states", " us", " nuclear", " fossil fuels", "
power"]\n          }\n          );\n          </script>\n\n\n\n</div>\n\n\n<div class="col-4
ml-md-auto dataset-card" id="finance">\n\n          <a href="finance/" class="dataset-
link float-left">\n\n          \n\n          </a>\n\n          <a href="finance/"
class="dataset-link">\n          <h3>Finance</h3>\n          </a>\n\n          <p>\n          The
Annual Survey of State Government Finances provides a comprehensive summary of
the annual survey findings for state governments, as well as data for individual
states.\n          <a href="finance/" class="dataset-link"><em>View</em></a>\n
</p>\n\n          <small>finance, government, expenditure, charges, money, united
states, us</small>\n\n          <script>\n          searcher.add(\n          {\n
\'name\': "finance",\n          \'body\': "The Annual Survey of State
Government Finances provides a comprehensive summary of the annual survey
findings for state governments, as well as data for individual states.",\n
\'description\': "The Annual Survey of State Government Finances provides a
comprehensive summary of the annual survey findings for state governments, as
well as data for individual states. The data contain detail of revenue by
source, expenditure by object and function, indebtedness by term, and assets by
purpose. This data goes back for almost two decades across every date. (1992 -
2019)",\n          \'tags\': ["finance", " government", " expenditure", "
charges", " money", " united states", " us"]\n          }\n          );\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="food">\n\n          <a href="food/" class="dataset-link float-left">\n\n          \n\n          </a>\n\n          <a href="food/" class="dataset-link">\n
<h3>Food</h3>\n          </a>\n\n          <p>\n          Statistics for various food items\n
<a href="food/" class="dataset-link"><em>View</em></a>\n          </p>\n\n
<small>food, vitamins, minerals, health, nutrition</small>\n\n          <script>\n
searcher.add(\n          {\n          \'name\': "food",\n          \'body\':
"Statistics for various food items",\n          \'description\': "The
following data comes from the United States Department of Agriculture\'s Food
Composition Database. It contains data for various types of food including the
amounts of different vitamins and minerals found in the foods as well as
macronutrient percentages. The food covered spans a large variety of foods from
butter to Campbell\'s soup. Much of the supplementary documentation for each
field comes directly from that pages\' Wikipedia article.",\n
\'tags\': ["food", " vitamins", " minerals", " health", " nutrition"]\n
}\n          );\n          </script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto
dataset-card" id="food access">\n\n          <a href="food_access/" class="dataset-
link float-left">\n\n          \n\n    <a
href="food_access/" class="dataset-link">\n    <h3>Food Access</h3>\n
</a>\n\n    <p>\n        Data about counties ability to access supermarkets,
supercenters, grocery stores, or other sources of healthy and affordable food.\n
<a href="food_access/" class="dataset-link"><em>View</em></a>\n    </p>\n\n
<small>counties, states, food, access, availability, supermarket,...</small>\n\n
<script>\n    searcher.add(\n        {\n            \'name\': "food_access",\n
\'body\': "Data about counties ability to access supermarkets, supercenters,
grocery stores, or other sources of healthy and affordable food.",\n
\'description\': "From the United States Department of Agriculture\'s Economic
Research Service, the dataset contains information about US county\'s ability to
access supermarkets, supercenters, grocery stores, or other sources of healthy
and affordable food. Most measures of how individuals and neighborhoods are able
to access food are based on the following indicators: - Accessibility to sources
of healthy food, as measured by distance to a store or by the number of stores
in an area. - Individual-level resources that may affect accessibility, such as
family income or vehicle availability. - Neighborhood-level indicators of
resources, such as the average income of the neighborhood and the availability
of public transportation.",\n            \'tags\': ["counties", " states", "
food", " access", " availability", " supermarket", " rural", " urban", "
population", " vehicles"]\n        }\n    );\n
</script>\n\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card"
id="global_development">\n\n    <a href="global_development/" class="dataset-
link float-left">\n\n        \n\n        </a>\n\n
<a href="global_development/" class="dataset-link">\n        <h3>Global
Development</h3>\n        </a>\n\n        <p>\n            Reports of country\'s development
over time\n            <a href="global_development/" class="dataset-
link"><em>View</em></a>\n        </p>\n\n        <small>world, countries, development,
health, population, urban,...</small>\n\n        <script>\n            searcher.add(\n
{\n                \'name\': "global_development",\n                \'body\': "Reports
of country\'s development over time",\n                \'description\': "The
following data contains records collected on different countries and geographic
locations from 1980 - 2013 from the World Bank. Included is different data about
urban development, agriculture and rural development, health, and
infrastructure.",\n                \'tags\': ["world", " countries", " development",
" health", " population", " urban", " rural", " agriculture", "
infrastructure"]\n            }\n        );\n        </script>\n\n\n</div>\n\n<div
class="col-4 ml-md-auto dataset-card" id="graduates">\n\n    <a
href="graduates/" class="dataset-link float-left">\n\n        \n\n        </a>\n\n        <a href="graduates/" class="dataset-
link">\n        <h3>Graduates</h3>\n        </a>\n\n        <p>\n            This library holds
data about employment of new graduates by major.\n            <a href="graduates/"
class="dataset-link"><em>View</em></a>\n        </p>\n\n        <small>graduation,
grads, majors, education, college, university,...</small>\n\n        <script>\n
searcher.add(\n            {\n                \'name\': "graduates",\n

```

```

\'body\': "This library holds data about employment of new graduates by
major.",\n          \'description\': "The data in this library comes from the
National Survey of Recent College Graduates. Included is information about
employment numbers, major information, and the earnings of different majors.
Many majors were not available before 2010, so their values have been recorded
as 0 (note that this may affect the averages shown in the bar charts).",\n
\'tags\': ["graduation", " grads", " majors", " education", " college", "
university", " higher education", " jobs", " careers"]\n          }\n          );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="health">\n\n      <a href="health/" class="dataset-link float-left">\n\n
\n\n      </a>\n\n      <a href="health/" class="dataset-link">\n
<h3>Health</h3>\n      </a>\n\n      <p>\n          This library holds data about
reports of diseases in America.\n          <a href="health/" class="dataset-
link"><em>View</em></a>\n      </p>\n\n      <small>health, diseases,
infection</small>\n\n      <script>\n          searcher.add(\n              {\n
\'name\': "health",\n              \'body\': "This library holds data about
reports of diseases in America.",\n              \'description\': "This library
holds data about reports of diseases in America.",\n              \'tags\':
["health", " diseases", " infection"]\n              }\n              );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="hospitals">\n\n      <a href="hospitals/" class="dataset-link float-
left">\n\n      \n\n      </a>\n\n      <a href="hospitals/"
class="dataset-link">\n      <h3>Hospitals</h3>\n      </a>\n\n      <p>\n          To
help consumers make informed decisions about health care, the Centers for
Medicare & Medicaid Services (CMS) collects data about the cost and quality
of care at over 4,000 Medicare-qualified hospitals.\n          <a
href="hospitals/" class="dataset-link"><em>View</em></a>\n      </p>\n\n
<small>hospitals, health care, medical, hospital costs, hospital
quality</small>\n\n      <script>\n          searcher.add(\n              {\n
\'name\': "hospitals",\n              \'body\': "To help consumers make informed
decisions about health care, the Centers for Medicare & Medicaid Services (CMS)
collects data about the cost and quality of care at over 4,000 Medicare-
qualified hospitals.",\n              \'description\': "The data set allows
consumers to directly compare across hospitals performance measure information
related to heart attack, emergency department care, preventive care, stroke
care, and other conditions. The data is part of an Administration-wide effort to
increase the availability and accessibility of information on quality,
utilization, and costs for effective, informed decision-making.",\n
\'tags\': ["hospitals", " health care", " medical", " hospital costs", "
hospital quality"]\n              }\n              );\n      </script>\n\n\n\n</div>\n\n\n<div
class="col-4 ml-md-auto dataset-card" id="hydropower">\n\n      <a
href="hydropower/" class="dataset-link float-left">\n\n      \n\n      </a>\n\n      <a href="hydropower/" class="dataset-
link">\n      <h3>Hydropower</h3>\n      </a>\n\n      <p>\n          Records about dams

```

```

in the United States such as location, dimensions, and project information\n
<a href="hydropower/" class="dataset-link"><em>View</em></a>\n    </p>\n\n
<small>hydropower, dams, lakes, rivers, waterways, irrigation,...</small>\n\n
<script>\n    searcher.add(\n        {\n            \'name\': "hydropower",\n            \'body\': "Records about dams in the United States such as location, dimensions,\nand project information",\n            \'description\': "Recent Federal policies\nand legislation focus on moving the nation towards a cleaner energy economy that\nincludes developing environmentally appropriate renewable energy projects\ninvolving solar, wind and waves, geothermal, biofuels, and hydropower. The 2010\nFederal Memorandum of Understanding for Hydropower and the Energy Policy Act of\n2005 direct the U.S. Department of Reclamation to evaluate development of new\nhydropower projects at Federally-owned facilities and upgrade or rehabilitate\nexisting hydropower generation facilities, as a contribution to the nation\'s\nclean energy goals. The following data describes the facilities currently being\noperated that contribute to nationwide renewable energy strategies.",\n            \'tags\': ["hydropower", " dams", " lakes", " rivers", " waterways", " irrigation", " human consumption", " industrial use", " aquaculture", " navigability", " energy", " electricity"]\n        }\n    );\n</script>\n\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card"\nid="ingredients">\n\n    <a href="ingredients/" class="dataset-link float-left">\n\n        \n\n        <a href="ingredients/" class="dataset-link">\n            <h3>Ingredients</h3>\n        </a>\n\n        <p>\n            Statistics for various food ingredients\n            <a href="ingredients/" class="dataset-link"><em>View</em></a>\n        </p>\n\n        <small>ingredients, vitamins, minerals, health, nutrition</small>\n\n        <script>\n            searcher.add(\n                {\n                    \'name\': "ingredients",\n                    \'body\': "Statistics for various food ingredients",\n                    \'description\': "The following data comes from the United States Department of\nAgriculture\'s Food Composition Database. It contains data for various types of\nfood ingredients including the amounts of different vitamins and minerals found\nin the foods as well as macronutrient percentages. The food covered spans a\nlarge variety of foods from butter to Campbell\'s soup. Much of the\nsupplementary documentation for each field comes directly from that pages\' Wikipedia article.",\n                    \'tags\': ["ingredients", " vitamins", " minerals", " health", " nutrition"]\n                }\n            );\n        </script>\n\n\n</div>\n\n<div class="col-4 ml-md-auto dataset-card"\nid="injuries">\n\n    <a href="injuries/" class="dataset-link float-left">\n\n        \n\n        <a href="injuries/" class="dataset-link">\n            <h3>Injuries</h3>\n        </a>\n\n        <p>\n            A sampling of work-related\n            injury and illness data from the US.\n            <a href="injuries/" class="dataset-link"><em>View</em></a>\n        </p>\n\n        <small>osha, united\nstates, us, usa, government, health, safety,...</small>\n\n        <script>\n            searcher.add(\n                {\n                    \'name\': "injuries",\n                    \'body\': "A sampling of work-related injury and illness data from the US.",\n                    \'description\': "The Occupational Safety and Health Administration (OSHA)

```

collected work-related injury and illness data from employers within specific industry and employment size specifications from 2002 through 2011. This data collection is called the OSHA Data Initiative or ODI. The data provided is used by OSHA to calculate establishment specific injury and illness incidence rates. This searchable database contains a table with the name, address, industry, and associated Total Case Rate (TCR), Days Away, Restricted, and Transfer (DART) case rate, and the Days Away From Work (DAFWII) case rate for the establishments that provided OSHA with valid data for calendar years 2002 through 2011. This data has been sampled down from its original size to 4%. In addition, the original dataset only has data from a small portion of all private sector establishments in the United States (80,000 out of 7.5 million total establishments). Therefore, these data are not representative of all businesses and general conclusions pertaining to all US business should not be overdrawn.

Data quality: While OSHA takes multiple steps to ensure the data collected is accurate, problems and errors invariably exist for a small percentage of establishments. OSHA does not believe the data for the establishments with the highest rates on this file are accurate in absolute terms. Efforts were made during the collection cycle to correct submission errors, however some remain unresolved. It would be a mistake to say establishments with the highest rates on this file are the \"most dangerous\" or \"worst\" establishments in the Nation.

Rate Calculation: An incidence rate of injuries and illnesses is computed from the following formula: (Number of injuries and illnesses X 200,000) / Employee hours worked = Incidence rate. The Total Case Rate includes all cases recorded on the OSHA Form 300 (Column G + Column H + Column I + Column J). The Days Away/Restricted/Transfer includes cases recorded in Column H + Column I. The Days Away includes cases recorded in Column H. For further information on injury and illness incidence rates, please visit the Bureau of Labor Statistics' webpage at <http://www.bls.gov/iif/osheval.htm>

State Participation: Not all state plan states participate in the ODI. The following states did not participate in the 2010 ODI (collection of CY 2009 data), establishment data is not available for these states: Alaska; Oregon; Puerto Rico; South Carolina; Washington; Wyoming.

```

</script>
</div>
<div class="col-4 ml-md-auto dataset-card"
id="labor">
  <a href="labor/" class="dataset-link float-left">
    
  </a>
  <a href="labor/" class="dataset-link">
    <h3>Labor</h3>
  </a>
  <p>
    Employment status of the civilian population by race, sex, and age.
  </p>
  <a href="labor/" class="dataset-link">
    <em>View</em>
  </a>
  <p>
    labor, race, age, sex, gender, america, usa, census,...
  </p>
  <script>
    searcher.add(
      {
        name: "labor",
        body: "Employment status of the civilian population by race, sex, and age.",
        description: "The Current Population Survey (CPS) is a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics. It provides a comprehensive

```

body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics.",\n

```

    \tags\': ["labor", " race", " age", " sex", "
gender", " america", " usa", " census", " employed", " unemployed", "
employability", " job", " work", " civilian", " black", " white", " asian", "
government"]\n    }\n    );\n    </script>\n\n\n</div>\n\n\n<div
class="col-4 ml-md-auto dataset-card" id="medal of honor">\n\n    <a
href="medal_of_honor/" class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a href="medal_of_honor/"
class="dataset-link">\n    <h3>Medal Of Honor</h3>\n    </a>\n\n    <p>\n
This dataset has records for the awarding of the United States Medal of Honor,
one of the military\'s highest honors.\n    <a href="medal_of_honor/"
class="dataset-link"><em>View</em></a>\n    </p>\n\n    <small>us, usa, united
states, military, combat, government, army,...</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "medal_of_honor",\n
\'body\': "This dataset has records for the awarding of the United States Medal
of Honor, one of the military\'s highest honors.",\n            \'description\':
"This dataset has records for the awarding of the United States Medal of Honor.
The Medal of Honor is the United States of America\'s highest military honor,
awarded for personal acts of valor above and beyond the call of duty. The medal
is awarded by the President of the United States in the name of the U.S.
Congress to U.S. military personnel only. There are three versions of the medal,
one for the Army, one for the Navy, and one for the Air Force.[5] Personnel of
the Marine Corps and Coast Guard receive the Navy version. The dataset was
collected from the official military site, and includes records about how the
medal was awarded and characteristics of the recipient. Unfortunately, because
of the nature of century-old record keeping, many of the records are incomplete.
While a very interesting dataset, it does have some missing data.",\n
\'tags\': ["us", " usa", " united states", " military", " combat", "
government", " army", " airforce", " marines", " navy", " fighting", " war", "
award", " medal", " honor", " glory", " death"]\n        }\n    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="music">\n\n    <a href="music/" class="dataset-link float-left">\n\n
\n\n    </a>\n\n    <a href="music/" class="dataset-link">\n
<h3>Music</h3>\n    </a>\n\n    <p>\n        The Music library is a compilation
of over one million contemporary songs and information about their audio
features and metadata.\n        <a href="music/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>music, songs, artists,
creativity, media</small>\n\n    <script>\n        searcher.add(\n            {\n
\'name\': "music",\n            \'body\': "The Music library is a compilation of
over one million contemporary songs and information about their audio features
and metadata.",\n            \'description\': "This library comes from the
Million Song Dataset, which used a company called the Echo Nest to derive data
points about one million popular contemporary songs. The Million Song Dataset is
a collaboration between the Echo Nest and LabROSA, a laboratory working towards

```

intelligent machine listening. The project was also funded in part by the National Science Foundation of America (NSF) to provide a large data set to evaluate research related to algorithms on a commercial size while promoting further research into the Music Information Retrieval field. The data contains standard information about the songs such as artist name, title, and year released. Additionally, the data contains more advanced information; for example, the length of the song, how many musical bars long the song is, and how long the fade in to the song was.",\n

```

    \tags\': ["music", " songs", "
artists", " creativity", " media"]\n
    }\n
    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="opioids">\n\n    <a href="opioids/" class="dataset-link float-left">\n\n
\n\n    </a>\n\n    <a href="opioids/" class="dataset-link">\n
<h3>Opioids</h3>\n    </a>\n\n    <p>\n        The data set shows the number and
rates of deaths due to opioid overdose.\n        <a href="opioids/"
class="dataset-link"><em>View</em></a>\n    </p>\n\n    <small>opioids, drug
deaths, heroin, cocaine, drug overdose</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "opioids",\n
\'body\': "The data set shows the number and rates of deaths due to opioid
overdose.",\n            \'description\': "All opioids are chemically related
and interact with opioid receptors on nerve cells in the body and brain. Opioid
pain relievers can be misused (taken in a different way or in a larger quantity
than prescribed, or taken without a doctor\'s prescription). Regular use - even
as prescribed by a doctor - can lead to dependence and, when misused, opioid
pain relievers can lead to addiction, overdose incidents, and deaths. The
National Institute on Drug Abuse collects and analyzes data about deaths from
opioid abuse. This data set reports on data from 1999-2019.",\n
\'tags\': ["opioids", " drug deaths", " heroin", " cocaine", " drug overdose"]\n
}\n    );\n    </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto
dataset-card" id="police shootings">\n\n    <a href="police_shootings/"
class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a href="police_shootings/"
class="dataset-link">\n    <h3>Police Shootings</h3>\n    </a>\n\n    <p>\n
The Washington Post is compiling a database of every fatal shooting in the
United States by a police officer in the line of duty since Jan. 1, 2015 by
culling local news reports, law enforcement websites and social media and by
monitoring independent databases. The Post conducted additional reporting in
many cases. \n        <a href="police_shootings/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>violence, crime, violent,
police, shootings, race, location, time</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "police_shootings",\n
\'body\': "The Washington Post is compiling a database of every fatal shooting
in the United States by a police officer in the line of duty since Jan. 1, 2015
by culling local news reports, law enforcement websites and social media and by
monitoring independent databases. The Post conducted additional reporting in
many cases.",\n            \'description\':

```



```

"https://www.washingtonpost.com/national/how-the-washington-post-is-examining-
police-shootings-in-the-united-
states/2016/07/07/d9c52238-43ad-11e6-8856-f26de2537a9d_story.html",\n
\'tags\': ["violence", " crime", " violent", " police", " shootings", " race", "
location", " time"]\n      }\n    );\n    </script>\n\n\n\n</div>\n\n\n<div
class="col-4 ml-md-auto dataset-card" id="publishers">\n\n    <a
href="publishers/" class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a href="publishers/" class="dataset-
link">\n    <h3>Publishers</h3>\n    </a>\n\n    <p>\n        E-book sales on
Amazon, including daily and total earnings for 54,000 titles.\n\n        <a
href="publishers/" class="dataset-link"><em>View</em></a>\n    </p>\n\n
<small>publishers, amazon, books, sales, genres, literature, english</small>\n\n
<script>\n    searcher.add(\n        {\n            \'name\': "publishers",\n
\'body\': "E-book sales on Amazon, including daily and total earnings for 54,000
titles.",\n            \'description\': "From a newspaper article about
analyzing amazon e-book sales by genre and publisher. Unfortunately, they do not
have information on the book\'s title or author. This collection includes 54,000
titles spanning across several genres and types of publishing companies,
practically every book on every Amazon bestseller list. Along with publisher
information, it also includes the book\'s overall Amazon Kindle store sales
ranking. This ranking is used to sort the books. Keep in mind that this data is
NOT time-oriented; it is a collection of a bunch of different books, not a book
over time.",\n            \'tags\': ["publishers", " amazon", " books", "
sales", " genres", " literature", " english"]\n        }\n    );\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card" id="real
estate">\n\n    <a href="real_estate/" class="dataset-link float-left">\n\n
\n\n    </a>\n\n    <a href="real_estate/"
class="dataset-link">\n    <h3>Real Estate</h3>\n    </a>\n\n    <p>\n
Real estate information in the United States, including inventory, building, and
customer data.\n        <a href="real_estate/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>real, estate, buildings,
government, building</small>\n\n    <script>\n    searcher.add(\n        {\n
\'name\': "real_estate",\n            \'body\': "Real estate information in the
United States, including inventory, building, and customer data.",\n            \'
description\': "Real Estate Across the United States (REXUS) is the primary
tool used by PBS to track and manage the government\'s real property assets and
to store inventory data, building data, customer data, and lease information.
STAR manages aspects of real property space management, including identification
of all building space and daily management of 22,000 assignments for all
property to its client Federal agencies. This data set contains PBS building
inventory that consists of both owned and leased buildings with active and
excess status.",\n            \'tags\': ["real", " estate", " buildings", "
government", " building"]\n        }\n    );\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="retail_services">\n\n    <a href="retail_services/" class="dataset-link

```

```

float-left">\n\n    \n\n    <a
href="retail_services/" class="dataset-link">\n    <h3>Retail Services</h3>\n
</a>\n\n    <p>\n        Monthly sales, inventories, and inventory/sales ratios
for retail trade and food services.\n        <a href="retail_services/"
class="dataset-link"><em>View</em></a>\n    </p>\n\n    <small>retail, services,
government, united states, usa, us, trade,...</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "retail_services",\n
\'body\': "Monthly sales, inventories, and inventory/sales ratios for retail
trade and food services.",\n            \'description\': "The Advance Monthly
and Monthly Retail Trade Surveys (MARTS and MRTS), the Annual Retail Trade
Survey (ARTS), and the Quarterly E-Commerce Report work together to produce the
most comprehensive data available on retail economic activity in the United
States. More detailed descriptions of these programs can be found by choosing
one of the links to the left. Regular quality control and verification takes
place between MARTS, MRTS, and ARTS annually and between these programs and the
Economic Census of Retail Trade every five years. Each year when annual data
become available, we compare and resolve differences between the data collected
on the monthly and annual surveys. We refer to this process as the monthly-to-
annual reconciliation. At the same time, we benchmark the monthly estimates
using results of the annual survey. ARTS estimates are then benchmarked to data
maintained by the Economic Census of Retail Trade. This process of benchmarking
retail data over all four programs ensures consistency in our estimates. The
Business Expenses Supplement is an addition to the 2007 Annual Retail Trade
Survey. Its purpose is to compile statistics on detailed business operating
expenses. The United States Code, Title 13, authorizes this program as part of
the Economic Census. This coverage was previously part of the predecessor
Business Expenses Survey. Detailed expenses are collected every five years.",\n
\'tags\': ["retail", " services", " government", " united states", " usa", "
us", " trade", " food", " services"]\n        }\n    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="school_scores">\n\n    <a href="school_scores/" class="dataset-link float-
left">\n\n    \n\n    <a
href="school_scores/" class="dataset-link">\n    <h3>School Scores</h3>\n
</a>\n\n    <p>\n        SAT scores for students across the United States,
organized by year and state.\n        <a href="school_scores/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>school, education, K-12,
kindergarten, learning, teaching, sat</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "school_scores",\n
\'body\': "SAT scores for students across the United States, organized by year
and state.",\n            \'description\': "SAT Scores from across the
country",\n            \'tags\': ["school", " education", " K-12", "
kindergarten", " learning", " teaching", " sat"]\n        }\n    );\n
</script>\n\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="skyscrapers">\n\n    <a href="skyscrapers/" class="dataset-link float-
left">\n\n    \n\n    </a>\n\n    <a
href="skyscrapers/" class="dataset-link">\n    <h3>Skyscrapers</h3>\n
</a>\n\n    <p>\n        Information about tall buildings constructed all over
the world.\n        <a href="skyscrapers/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>buildings, construction,
building, skyscrapers,...</small>\n\n    <script>\n        searcher.add(\n
{\n            \'name\': "skyscrapers",\n            \'body\': "Information
about tall buildings constructed all over the world.",\n
\'description\': "Collected statistics about the tallest buildings all over the
world, including the materials used to develop them, the purpose of the
buildings, and more.",\n            \'tags\': ["buildings", " construction", "
building", " skyscrapers", " international", " tall", " height", " tower"]\n
}\n        );\n    </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto
dataset-card" id="slavery">\n\n    <a href="slavery/" class="dataset-link
float-left">\n\n        \n\n        </a>\n\n        <a href="slavery/"
class="dataset-link">\n            <h3>Slavery</h3>\n        </a>\n\n        <p>\n            This
dataset has financial records of New Orleans slave sales, 1856-1861. There are
almost 16,000 sales recorded in this dataset.\n            <a href="slavery/"
class="dataset-link"><em>View</em></a>\n        </p>\n\n        <small>slavery, slave,
slaves, buyer, seller, origin, history, economics</small>\n\n        <script>\n
searcher.add(\n            {\n                \'name\': "slavery",\n
\'body\': "This dataset has financial records of New Orleans slave sales,
1856-1861. There are almost 16,000 sales recorded in this dataset.",\n
\'description\': "This dataset has financial records of New Orleans slave sales,
1856-1861. There are almost 16,000 sales recorded in this dataset.",\n
\'tags\': ["slavery", " slave", " slaves", " buyer", " seller", " origin", "
history", " economics"]\n            }\n        );\n    </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="state_crime">\n\n    <a href="state_crime/" class="dataset-link float-
left">\n\n        \n\n        </a>\n\n        <a
href="state_crime/" class="dataset-link">\n            <h3>State Crime</h3>\n
</a>\n\n        <p>\n            Records about the crime rates and totals for US states
over time.\n            <a href="state_crime/" class="dataset-
link"><em>View</em></a>\n        </p>\n\n        <small>crime, burglary, larceny, motor,
property, violent, assault,...</small>\n\n        <script>\n            searcher.add(\n
{\n                \'name\': "state_crime",\n                \'body\': "Records about
the crime rates and totals for US states over time.",\n
\'description\': "From the Unified Crime Reporting Statistics and under the
collaboration of the U.S. Department of Justice and the Federal Bureau of
Investigation information crime statistics are available for public review. The
following data set has information on the crime rates and totals for states
across the United States for a wide range of years. The crime reports are
divided into two main categories: property and violent crime. Property crime
refers to burglary, larceny, and motor related crime while violent crime refers
to assault, murder, rape, and robbery. These reports go from 1960 to 2019.",\n

```

```

\'tags\': ["crime", " burglary", " larceny", " motor", " property", " violent",
" assault", " murder", " rape", " robbery"]\n          }\n        );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="state_demographics">\n\n    <a href="state_demographics/" class="dataset-
link float-left">\n\n        \n\n        </a>\n\n    <a href="state_demographics/" class="dataset-link">\n        <h3>State
Demographics</h3>\n        </a>\n\n        <p>\n            Demographic information for the
states in the United States.\n            <a href="state_demographics/"
class="dataset-link"><em>View</em></a>\n            </p>\n\n        <small>demographics,
population, age, ethnicity, race, language,...</small>\n\n        <script>\n
searcher.add(\n            {\n                \'name\': "state_demographics",\n
\'body\': "Demographic information for the states in the United States.",\n
\'description\': "The following data set is summarized information obtained
about states in the United States from 2015 through 2019 through the United
States Census Bureau. However, this data is not available for these years; just
the summarized data as of 2019. Information described in the data includes the
age distributions, the education levels, employment statistics, ethnicity
percents, household information, income, and other miscellaneous statistics.
(Values are denoted as -1, if the data is not available)",\n
\'tags\': ["demographics", " population", " age", " ethnicity", " race", "
language", " education", " foreign", " households", " income", " employment", "
unemployment", " sales", " land area", " states"]\n                }\n            );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="state_fragility">\n\n    <a href="state_fragility/" class="dataset-link
float-left">\n\n        \n\n        </a>\n\n    <a
href="state_fragility/" class="dataset-link">\n        <h3>State Fragility</h3>\n
</a>\n\n    <p>\n        The State Fragility Index scores for countries around
the world from 1995 to 2018\n        <a href="state_fragility/" class="dataset-
link"><em>View</em></a>\n        </p>\n\n    <small>world, countries, security,
politics, economy, society,...</small>\n\n    <script>\n        searcher.add(\n
{\n            \'name\': "state_fragility",\n            \'body\': "The State
Fragility Index scores for countries around the world from 1995 to 2018",\n
\'description\': "The Center for Systemic Peace was founded in 1997 to engage in
global systems analysis to minimize the effects of political violence in the
world as a whole. The following data set shows the state fragility for countries
with a population greater than 500,000 in 2013. The State Fragility Index scores
countries on two main categories, effectiveness and legitimacy. These are then
broken down into four dimensions: Security, Political, Economic, and Social. The
State Fragility Index score is the sum of the individual country\'s
effectiveness score and their legitimacy score. Each of these scores are the
summation of the four dimensions within the category.",\n            \'tags\':
["world", " countries", " security", " politics", " economy", " society", "
effectiveness", " legitimacy"]\n            }\n        );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="suicide_attacks">\n\n    <a href="suicide_attacks/" class="dataset-link

```

```
float-left">\n\n    \n\n    </a>\n\n    <a
href="suicide_attacks/" class="dataset-link">\n    <h3>Suicide Attacks</h3>\n
</a>\n\n    <p>\n        This dataset has about suicide attacks all over the
world for the past three decades.\n        <a href="suicide_attacks/"
class="dataset-link"><em>View</em></a>\n    </p>\n\n    <small>global, attack,
damage, kill, hurt, destroy, pain,...</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "suicide_attacks",\n
\'body\': "This dataset has about suicide attacks all over the world for the
past three decades.",\n            \'description\': "The Chicago Project on
Security and Terrorism (CPOST) maintains a searchable database on all suicide
attacks from 1982 through October 2020. The database includes information about
the location of attacks, the target type, the weapon used, and systematic
information on the demographic and general biographical characteristics of
suicide attackers. The database expands the breadth of the data available in
English using native language sources (e.g., Arabic, Hebrew, Russian, Tamil)
that are likely to have the most extensive relevant information. The current
CPOST-SAD release contains the universe of suicide attacks from 1982 through
September 2015, a total of 4,814 attacks in over 40 countries. The CPOST Suicide
Attack Database (CPOST-SAD) contains data on the universe of suicide attacks.
For each attack, the database includes information about the geographic
location, target classification, and weapon(s) used, as well as systematic
information on the demographic and general biographical characteristics of
suicide attackers. More information about the coding and methodology is
available on the CPOST website below. Note that multiple attackers or targets
are collapsed into a single record for simplicity\'s sake.",\n
\'tags\': ["global", " attack", " damage", " kill", " hurt", " destroy", "
pain", " suffering", " terrorism", " terrorism", " religion", " warfare", " war", "
car", " bomb", " gun", " shooting"]\n        }\n    );\n
</script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="supreme court">\n\n    <a href="supreme_court/" class="dataset-link float-
left">\n\n    \n\n    </a>\n\n    <a
href="supreme_court/" class="dataset-link">\n    <h3>Supreme Court</h3>\n
</a>\n\n    <p>\n        This dataset contains information about supreme court
cases for the several decades.\n        <a href="supreme_court/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>us, usa, united states,
government, law, legal, court, case,...</small>\n\n    <script>\n
searcher.add(\n        {\n            \'name\': "supreme_court",\n
\'body\': "This dataset contains information about supreme court cases for the
several decades.",\n            \'description\': "The U.S. Supreme Court
Database traces its history back about two decades ago, when Harold J. Spaeth
asked the National Science Foundation to fund a database that would be so rich
in content that multiple users - even those with vastly distinct projects and
purposes in mind - could draw on it. Professor Spaeth\'s goal was at once
refreshingly simple and extremely ambitious: to produce a database that would
include and classify every single vote by a Supreme Court justice in all argued
```

cases over a five-decade period. After securing the funding, Spaeth collected and coded the data, performed reliability checks, and eventually amassed the Database. In the late 1980s, he made it (and the documentation necessary to use it) publicly available. Since then, Professor Spaeth has not only updated it each term; he has also continued to perform reliability analyses, thereby ensuring its integrity with each release, and added new variables. Today's version of the Database houses 247 pieces of information for each case, roughly broken down into six categories: (1) identification variables (e.g., citations and docket numbers); (2) background variables (e.g., how the Court took jurisdiction, origin and source of the case, the reason the Court agreed to decide it); (3) chronological variables (e.g., the date of decision, term of Court, natural court); (4) substantive variables (e.g., legal provisions, issues, direction of decision); (5) outcome variables (e.g., disposition of the case, winning party, formal alteration of precedent, declaration of unconstitutionality); and (6) voting and opinion variables (e.g., how the individual justices voted, their opinions and interagreements). Case Centered data provides case level information; i.e., each row in the database corresponds to a dispute. These data do not contain specific justice vote information. A tremendously helpful resource for learning more about the dataset is the codebook: <http://supremecourtdatabase.org/documentation.php?s=1>.

'tags': ["us", "usa", "united states", "government", "law", "legal", "court", "case", "supreme", "judge", "barrister", "lawyer", "federal", "national"]

};

</script>

<div class="col-4 ml-md-auto dataset-card" id="tate">

<h3>Tate</h3>

<p> This dataset is about the Tate art collection, with metadata about paintings, drawings, sculptures, and more.

View

<p>

<small>artwork, art, sculpture, tate</small>

<script>

searcher.add({

'name': "tate",

'body': "This dataset is about the Tate art collection, with metadata about paintings, drawings, sculptures, and more.",

'description': "Tate holds the national collection of British art from 1500 to the present day and international modern and contemporary art. The collection embraces all media, from painting, drawing, sculpture and prints to photography, video and film, installation and performance. This dataset is not ordered in any particular way. This dataset has metadata about many pieces from their collection.",

'tags': ["artwork", "art", "sculpture", "tate"]

});

</script>

<div class="col-4 ml-md-auto dataset-card" id="video games">

<h3>Video Games</h3>

<p> This library has data about video games, including their play time and sales.

<a href="video_games/" class="dataset-

```

link"><em>View</em></a>\n    </p>\n\n    <small>video games, games, publishers,
console, gaming</small>\n\n    <script>\n        searcher.add(\n            {\n
\'name\': "video_games",\n                \'body\': "This library has data about
video games, including their play time and sales.",\n
\'description\': "This dataset requires special permission to use. Originally
collected by Dr. Joe Cox, this dataset has information about the sales and
playtime of over a thousand video games released between 2004 and 2010. The
playtime information was collected from crowd-sourced data on \\"How Long to
Beat\\". Some more information can be found here.",\n                \'tags\':
["video games", " games", " publishers", " console", " gaming"]\n            }\n
); \n    </script>\n\n\n</div>\n\n\n<div class="col-4 ml-md-auto dataset-card"
id="weather">\n\n    <a href="weather/" class="dataset-link float-left">\n\n
\n\n    </a>\n\n    <a href="weather/" class="dataset-link">\n
<h3>Weather</h3>\n    </a>\n\n    <p>\n        Weekly weather records in 2016
across the U.S.\n        <a href="weather/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>weather, rain, snow, sleet, fog,
temperature, wind, climate,...</small>\n\n    <script>\n        searcher.add(\n
{\n            \'name\': "weather",\n                \'body\': "Weekly weather
records in 2016 across the U.S.",\n                \'description\': "Under the
National Oceanic and Atmospheric Administration, the National Weather Service
provides daily weather reports for cities across the county. This is done
through the use of 122 different Weather Forecast Offices throughout the country.
These WFOs are responsible for the daily weather reports for serveral cities
throughout their region of coverage. This data set takes the information from
these WFO reports for cities across the country and summarizes it at the weekly
level for all of 2016.",\n                \'tags\': ["weather", " rain", " snow", "
sleet", " fog", " temperature", " wind", " climate", " environment", "
geology"]\n            }\n        ); \n    </script>\n\n\n</div>\n\n\n<div class="col-4
ml-md-auto dataset-card" id="wind turbines">\n\n    <a href="wind_turbines/"
class="dataset-link float-left">\n\n    \n\n    </a>\n\n    <a href="wind_turbines/" class="dataset-
link">\n    <h3>Wind Turbines</h3>\n    </a>\n\n    <p>\n        This dataset
provides locations and technical specifications of wind turbines in the United
States, almost all of which are utility-scale. Utility-scale turbines are ones
that generate power and feed it into the grid, supplying a utility with energy.
They are usually much larger than turbines that would feed a homeowner or
business.\n        <a href="wind_turbines/" class="dataset-
link"><em>View</em></a>\n    </p>\n\n    <small>energy, wind, power, air,
turbine, utility, location, time</small>\n\n    <script>\n        searcher.add(\n
{\n            \'name\': "wind_turbines",\n                \'body\': "This dataset
provides locations and technical specifications of wind turbines in the United
States, almost all of which are utility-scale. Utility-scale turbines are ones
that generate power and feed it into the grid, supplying a utility with energy.
They are usually much larger than turbines that would feed a homeowner or
business.",\n                \'description\': "The Open PV Project is a

```

collaborative effort between government, industry, and the public that continues to compile a database of available public data for photovoltaic (PV) installation data for the United States. Data for the project are voluntarily contributed from a variety of sources including solar incentive programs, utilities, installers, and the general public. This database serves as a web-based resource for users to easily explore and understand the current and past trends of the US PV industry. The data collected is actively maintained by the contributors and are always changing to provide an evolving, up-to-date snapshot of the US solar power market."

```

\tags\': ["energy", " wind", "
power", " air", " turbine", " utility", " location", " time"]\n
});\n
</script>\n\n\n</div>\n\n</div>\n\n<hr />\n\n\n
</div>\n\n
<footer id="footer" class="site-footer">\n
<div class="copyright">\n
\n
<p>&copy; 2022 CORGIS Datasets Project. Project by Austin Cory Bart, Dennis
Kafura, Clifford A. Shaffer, Javier Tibau, Luke Gusukuma, Eli Tilevich.</p>\n
\n
</div>\n</footer>\n\n
<script
src="https://use.fontawesome.com/releases/v5.0.12/js/all.js"></script>\n<script
src="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/js/bootstrap.bundle.min.
js"></script>\n
<script type="text/javascript">\n
$(document).ready(function() {\n
    $("#search-bar").on(\'input
propertychange\', function() {\n
        let needle = $(this).val();\n
if (needle === "") {\n
        $(".dataset-card").show();\n
}
else {\n
        let results = searcher.search(needle);\n
$(".dataset-card").hide();\n
        results.forEach(function (data) {\n
$("#"+data.ref).show();\n
        });\n
        });\n
});\n
</script>\n\n
</body>\n
\n</html>'

```

```
[4]: print(response.text)
```

```

<!DOCTYPE html>
<html lang="en-US">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-
to-fit=no">

```

```

<title>CORGIS Datasets Project</title>
<meta name="description" content="

```



```

    Github Pages for CORGIS Datasets Project">
    <link rel="canonical" href="https://corgis-edu.github.io//corgis/csv/">

    <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css">
    <link rel="stylesheet"
href="https://code.jquery.com/ui/1.12.1/themes/base/jquery-ui.css">
    <link rel="stylesheet" href="/corgis/assets/css/main.css">


    <script src="/corgis/assets/js/jquery-3.3.1.min.js"></script>
    <script src="https://code.jquery.com/ui/1.12.1/jquery-ui.min.js"
        integrity="sha256-VazP97ZCwtekAsvgPBSUwPFKdrwD3unUfSGVYrahUqU="
        crossorigin="anonymous"></script>

    <script src="https://cdnjs.cloudflare.com/ajax/libs/knockout/3.5.0/knockout-
debug.js"
        integrity="sha256-M5ZomNNnrnEB2WjSbnty5GWGqq6UuAAVNnWECisgEis="
        crossorigin="anonymous"></script>

    <link rel="stylesheet" href="/corgis/assets/css/multiselect.css">
    <link rel="stylesheet" href="/corgis/assets/css/bootstrap-tagsinput.css">

    <script src="/corgis/assets/js/knockout-3.5.0.js"></script>
    <script src="/corgis/assets/js/d3.min.js"></script>
    <script src="/corgis/assets/js/jquery.multi-select.js"></script>
    <script src="/corgis/assets/js/bootstrap-tagsinput.js"></script>
    <script src="/corgis/assets/js/typeahead.bundle.js"></script>
    <script src="/corgis/assets/js/jquery.quicksearch.js"></script>
    <script src="/corgis/assets/js/jquery.multi-select.js"></script>
    <script src="/corgis/assets/js/lunr.js"></script>

</head>
    <body class="layout--default "
    >
    <div class="container-fluid">

    <script>
let searcher = lunr(function() {
    this.field('name', 10);
    this.field('tags', 100);

```

```

        this.field('body');
        this.field('description');
        this.ref('name');
    });
</script>

<p><a href="/corgis/" class="site-logo" rel="home" title="CORGIS Datasets
Project">
    
</a></p>

<h1 id="csv-datasets">CSV Datasets</h1>

<p class="lead">CORGIS: The Collection of Really Great, Interesting, Situated
Datasets</p>

<p><span class="text-muted">By Austin Cory Bart, Ryan Whitcomb, Jason Riddle,
Omar Saleem, Dr. Eli Tilevich, Dr. Clifford A. Shaffer, Dr. Dennis
Kafura</span></p>

<hr />

<form>
<div class="form-group row search-bar">
    <label for="search-bar" class="col-sm-1 col-form-label">Search:</label>
    <div class="col-sm-11">
        <input type="text" class="form-control" id="search-bar"
placeholder="Keyword or phrase" />
    </div>
</div>
</form>

<hr />

<div class="row">

<div class="col-4 ml-md-auto dataset-card" id="aids">

    <a href="aids/" class="dataset-link float-left">

    </a>

```

```

<a href="aids/" class="dataset-link">
<h3>Aids</h3>
</a>

<p>
  Records of AIDS related statistics from several countries.
  <a href="aids/" class="dataset-link"><em>View</em></a>
</p>

<small>aids, death, disease, hiv, orphans, health, countries,...</small>

<script>
searcher.add(
  {
    'name': "aids",
    'body': "Records of AIDS related statistics from several
countries.",
    'description': "The UNAIDS Organization is an entity of the United
Nations that looks to reduce the transmission of AIDS and provide resources to
those currently affected by the disease. The following data set contains
information on the number of those affected by the disease, new cases of the
disease being reported, and AIDS-related deaths for a large set of countries
over the course of 1990 - 2015.",
    'tags': ["aids", " death", " disease", " hiv", " orphans", "
health", " countries", " world", " gender", " united nations", " un"]
  }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="airlines">

```

```

  <a href="airlines/" class="dataset-link float-left">

  </a>

  <a href="airlines/" class="dataset-link">
  <h3>Airlines</h3>
  </a>

  <p>

```

```

        Information about flight delays in major airports since 2003.
        <a href="airlines/" class="dataset-link"><em>View</em></a>
    </p>

    <small>airplane, airports, travel, plane, air, flights, delays,...</small>

    <script>
    searcher.add(
        {
            'name': "airlines",
            'body': "Information about flight delays in major airports since
2003.",
            'description': "This dataset is all about flights in the united
states, including information about the number, length, and type of delays. The
data is reported for individual months at every major airport for every carrier.
Additional information is available:
http://www.rita.dot.gov/bts/help/aviation/html/understanding.html",
            'tags': ["airplane", " airports", " travel", " plane", " air", "
flights", " delays", " national", " united states", " transportation"]
        }
    );
    </script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="billionaires">

    <a href="billionaires/" class="dataset-link float-left">

    </a>

    <a href="billionaires/" class="dataset-link">
    <h3>Billionaires</h3>
    </a>

    <p>
        Information about over 2000 billionaires from around the world.
        <a href="billionaires/" class="dataset-link"><em>View</em></a>
    </p>

    <small>money, rich, wealthy, people, person, billionaire</small>

```

```

<script>
searcher.add(
  {
    'name': "billionaires",
    'body': "Information about over 2000 billionaires from around the
world.",
    'description': "Researchers have compiled a multi-decade database of
the super-rich. Building off the Forbes World's Billionaires lists from
1996-2014, scholars at Peterson Institute for International Economics have added
a couple dozen more variables about each billionaire - including whether they
were self-made or inherited their wealth. (Roughly half of European billionaires
and one-third of U.S. billionaires got a significant financial boost from
family, the authors estimate.)",
    'tags': ["money", " rich", " wealthy", " people", " person", "
billionaire"]
  }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="broadway">

```

```

  <a href="broadway/" class="dataset-link float-left">

```

```

```

```

  </a>

```

```

  <a href="broadway/" class="dataset-link">

```

```

    <h3>Broadway</h3>

```

```

  </a>

```

```

  <p>

```

```

    This library holds data about Broadway shows, such as tickets sold.

```

```

    <a href="broadway/" class="dataset-link"><em>View</em></a>

```

```

  </p>

```

```

  <small>broadway, musical, theatre, tickets</small>

```

```

  <script>

```

```

    searcher.add(

```

```

      {

```

```

        'name': "broadway",

```

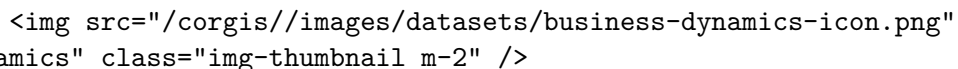
```

        'body': "This library holds data about Broadway shows, such as
tickets sold.",
        'description': "This library holds data about over Broadway shows,
grouped over weeklong periods. Only shows that reported capacity were included,
so the dataset stretches back to the 1990s. The dataset is made available by the
Broadway League (the national trade association for the Broadway industry), and
you can view the data online at http://www.broadwayleague.com/.",
        'tags': ["broadway", " musical", " theatre", " tickets"]
    }
);
</script>

```

</div>

<div class="col-4 ml-md-auto dataset-card" id="business dynamics">

 Business Dynamics" class="img-thumbnail m-2" />

Business Dynamics</h3>

<p>

The Business Dynamics Statistics (BDS) includes measures of establishment openings and closings, firm startups, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics for the US.

[View](business_dynamics/)

</p>

<small>government, united states, us, usa, business, businesses,...</small>

<script>

searcher.add(

{

 'name': "business_dynamics",

 'body': "The Business Dynamics Statistics (BDS) includes measures of establishment openings and closings, firm startups, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics for the US.",

'description': "The Business Dynamics Statistics (BDS) includes measures of establishment openings and closings, firm startups, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics. The U.S. economy is comprised of over 6 million establishments with paid employees. The population of these businesses is constantly churning -- some businesses grow, others decline and yet others close. New businesses are constantly replenishing this pool. The BDS series provide annual statistics on gross job gains and losses for the entire economy and by industrial sector, state, and MSA. These data track changes in employment at the establishment level, and thus provide a picture of the dynamics underlying aggregate net employment growth. There is a longstanding interest in the contribution of small businesses to job and productivity growth in the U.S. Some recent research suggests that it is business age rather than size that is the critical factor. The BDS permits exploring the respective contributions of both firm age and size. BDS is based on data going back through 1976. This allows business dynamics to be tracked, measured and analyzed for young firms in their first critical years as well as for more mature firms including those that are in the process of reinventing themselves in an ever changing economic environment. If you need help understanding the terms used, check out these definitions.",

'tags': ["government", " united states", " us", " usa", " business",
" businesses", " firms", " establishments", " jobs", " census"]
}

);

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="cancer">

<h3>Cancer</h3>

<p>

Cumulative cancer deaths for the period 2007-2013 are reported for each U.S. state.

View

```

</p>

<small>cancer, cancer deaths, medical, health</small>

<script>
searcher.add(
  {
    'name': "cancer",
    'body': "Cumulative cancer deaths for the period 2007-2013 are
reported for each U.S. state.",
    'description': "Information about the rates of cancer deaths in each
state is reported. The data shows the total rate as well as rates based on sex,
age, and race. Rates are also shown for three specific kinds of cancer: breast
cancer, colorectal cancer, and lung cancer.",
    'tags': ["cancer", " cancer deaths", " medical", " health"]
  }
);
</script>

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="cars">

  <a href="cars/" class="dataset-link float-left">

  </a>

  <a href="cars/" class="dataset-link">
  <h3>Cars</h3>
  </a>

  <p>
    This is a dataset about cars and how much fuel they use.
    <a href="cars/" class="dataset-link"><em>View</em></a>
  </p>

  <small>cars, vehicles, fuel</small>

  <script>
searcher.add(
  {
    'name': "cars",

```



```

        'body': "This is a dataset about cars and how much fuel they use.",
        'description': "This is a dataset about cars and how much fuel they
use.",
        'tags': ["cars", " vehicles", " fuel"]
    }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="classics">

```

```

    <a href="classics/" class="dataset-link float-left">

```

```

```

```

    </a>

```

```

    <a href="classics/" class="dataset-link">

```

```

        <h3>Classics</h3>

```

```

    </a>

```

```

    <p>

```

```

        Records and computed statistics about the top 1000 books on Project
Gutenberg.

```

```

        <a href="classics/" class="dataset-link"><em>View</em></a>

```

```

    </p>

```

```

    <small>classics, books, texts, text, book, classic, english,...</small>

```

```

    <script>

```

```

        searcher.add(

```

```

            {

```

```

                'name': "classics",

```

```

                'body': "Records and computed statistics about the top 1000 books on
Project Gutenberg.",

```

```

                'description': "Project Gutenberg (PG) is a volunteer effort to
digitize and archive cultural works, to 'encourage the creation and distribution
of eBooks'. It was founded in 1971 by Michael S. Hart and is the oldest digital
library. This dataset is a collection of the top 1000 most popular books on
Project Gutenberg, as determined by downloads. Each book has information about
its authorship, publication date, congressional classication, and a few other
fields. It also has some simple, computed statistics based on common metrics
such as sentiment analysis, Flesch Kincaid Reading level, and average sentence

```

```

length.",
    'tags': ["classics", " books", " texts", " text", " book", "
classic", " english", " shakespeare", " literature", " novel", " language", "
composition", " writing", " author", " publication", " words"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="construction permits">

    <a href="construction_permits/" class="dataset-link float-left">

    </a>

    <a href="construction_permits/" class="dataset-link">
<h3>Construction Permits</h3>
</a>

    <p>
        This dataset provides data on the number and valuation of new housing
units authorized by building permits.
        <a href="construction_permits/" class="dataset-link"><em>View</em></a>
    </p>

    <small>construction, permits, us, united states, america,...</small>

    <script>
searcher.add(
    {
        'name': "construction_permits",
        'body': "This dataset provides data on the number and valuation of
new housing units authorized by building permits.",
        'description': "This dataset provides data on the number and
valuation of new housing units authorized by building permits. Data are
available monthly at the national, regional, divisional, and state level. The
data are from the Building Permits Survey. The United States Code, Title 13,
authorizes this survey, provides for voluntary responses, and provides an
exception to confidentiality for public records.",
        'tags': ["construction", " permits", " us", " united states", "
america", " government", " buildings", " private", " units", " valuation", "

```

```
housing"]
    }
);
</script>
```

```
</div>
```

```
<div class="col-4 ml-md-auto dataset-card" id="construction spending">
```

```
    <a href="construction_spending/" class="dataset-link float-left">
```

```
        
```

```
    </a>
```

```
    <a href="construction_spending/" class="dataset-link">
```

```
        <h3>Construction Spending</h3>
```

```
    </a>
```

```
    <p>
```

```
        Estimates of the total dollar value of construction work done in the
    U.S.
```

```
        <a href="construction_spending/" class="dataset-link"><em>View</em></a>
```

```
    </p>
```

```
    <small>construction, spending, us, united states, america,...</small>
```

```
    <script>
```

```
    searcher.add(
        {
            'name': "construction_spending",
            'body': "Estimates of the total dollar value of construction work
done in the U.S.",
            'description': "The Value of Construction Put in Place Survey (VIP)
provides monthly estimates of the total dollar value of construction work done
in the U.S. The United States Code, Title 13, authorizes this program. The
survey covers construction work done each month on new structures or
improvements to existing structures for private and public sectors. Data
estimates include the cost of labor and materials, cost of architectural and
engineering work, overhead costs, interest and taxes paid during construction,
and contractor's profits. Data collection and estimation activities begin on the
first day after the reference month and continue for about three weeks. Reported
data and estimates are for activity taking place during the previous calendar
month. The survey has been conducted monthly since 1964. Why isn't the 'total
```

construction' the sum of the other categories? It is! the trick is that the 'residential' and \"nonresidential\" categories are supposed to be below the 'total construction' level, and the other categories are below those two categories. This example of the data should make it more clear:
<http://www.census.gov/construction/c30/pdf/tot.pdf>. More information about the nature of the construction projects is also available.",

```
    'tags': ["construction", " spending", " us", " united states", "
america", " government", " buildings", " residential", " non-residential", "
private", " public"]
  }
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="county demographics">

<h3>County Demographics</h3>

<p>

Demographic information for counties in the United States.

View

</p>

<small>demographics, population, age, ethnicity, race, language,...</small>

<script>

searcher.add(

{

'name': "county_demographics",

'body': "Demographic information for counties in the United

States.",

'description': "The following data set is information obtained about

counties in the United States from 2010 through 2019 through the United States

Census Bureau. Information described in the data includes the age distributions,

the education levels, employment statistics, ethnicity percents, houseold

information, income, and other miscellaneous statistics. (Values are denoted as -1, if the data is not available)",

```
    'tags': ["demographics", " population", " age", " ethnicity", "
race", " language", " education", " foreign", " households", " income", "
employment", " unemployment", " sales", " land area", " counties"]
  }
```

```
);
```

```
</script>
```

```
</div>
```

```
<div class="col-4 ml-md-auto dataset-card" id="covid">
```

```
  <a href="covid/" class="dataset-link float-left">
```

```
    
```

```
  </a>
```

```
  <a href="covid/" class="dataset-link">
```

```
    <h3>Covid</h3>
```

```
  </a>
```

```
  <p>
```

Since the beginning of the coronavirus pandemic, the Epidemic Intelligence team of the European Center for Disease Control and Prevention (ECDC) has been collecting on daily basis the number of COVID-19 cases and deaths, based on reports from health authorities worldwide. This helps to monitor and interpret the dynamics of the COVID-19 pandemic not only in the European Union (EU), the European Economic Area (EEA), but also worldwide. Every day a team of epidemiologists screens up to 500 relevant sources to collect the latest figures. The data screening is followed by ECDC standard epidemic intelligence process for which every single data entry is validated and documented in an ECDC database.

```
    <a href="covid/" class="dataset-link"><em>View</em></a>
```

```
  </p>
```

```
  <small>Covid, Covid-19, pandemic, infection, world health</small>
```

```
  <script>
```

```
    searcher.add(
```

```
      {
```

```
        'name': "covid",
```

```
        'body': "Since the beginning of the coronavirus pandemic, the
```

Epidemic Intelligence team of the European Center for Disease Control and Prevention (ECDC) has been collecting on daily basis the number of COVID-19 cases and deaths, based on reports from health authorities worldwide. This helps to monitor and interpret the dynamics of the COVID-19 pandemic not only in the European Union (EU), the European Economic Area (EEA), but also worldwide. Every day a team of epidemiologists screens up to 500 relevant sources to collect the latest figures. The data screening is followed by ECDC standard epidemic intelligence process for which every single data entry is validated and documented in an ECDC database.",

'description': "The data set contains daily reports of Covid-19 cases and deaths in countries worldwide. The data also shows the country's population and the number of cases per 100,000 people on a rolling 14 day average.",

'tags': ["Covid", " Covid-19", " pandemic", " infection", " world health"]
}
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="drugs">

<h3>Drugs</h3>

<p>

This dataset is about substance abuse (cigarettes, marijuana, cocaine, alcohol) among different age groups and states.

View

</p>

<small>drugs, substances, drug usage</small>

<script>

searcher.add(

{

```

        'name': "drugs",
        'body': "This dataset is about substance abuse (cigarettes,
marijuana, cocaine, alcohol) among different age groups and states.",
        'description': "This dataset is about substance abuse (cigarettes,
marijuana, cocaine, alcohol) among different age groups and states. Data was
collected from individual states as part of the NSDUH study. The data ranges
from 2002 to 2018. Both totals (in thousands of people) and rates (as a
percentage of the population) are given.",
        'tags': ["drugs", " substances", " drug usage"]
    }
);
</script>

```

</div>

<div class="col-4 ml-md-auto dataset-card" id="earthquakes">

<h3>Earthquakes</h3>

<p>

Records from different earthquake occurrences across the world.

View

</p>

<small>earthquakes, nature, disaster, magnitude, richter scale,...</small>

<script>

searcher.add(

{

'name': "earthquakes",

'body': "Records from different earthquake occurrences across the
world.",

'description': "Earthquake records from around the world collected
from the United States Geological Survey. Important details about the earthquake
such as distance, gap, magnitude, depth and significance are included to
properly describe the earthquake. Additionally, data about exact geological

coordinates and a relative description of the earthquake's location is included. The earthquakes collected are from the past month.",

```
    'tags': ["earthquakes", " nature", " disaster", " magnitude", " richter scale", " latitude", " longitude"]
  }
```

```
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="election">

<h3>Election</h3>

<p>

A breakdown of how each county voted in the 2016 Presidential primaries.

View

</p>

<small>politics, election, votes, counties, candidate</small>

<script>

searcher.add(

{

'name': "election",

'body': "A breakdown of how each county voted in the 2016

Presidential primaries.",

'description': "Data about the results was collected from CNN's 2016 election center. Each county has the breakdown of how many votes each candidate recieved in that county and what percentage of the total vote from that county that the candidate recieved. The candidate's political party is also represented.",

'tags': ["politics", " election", " votes", " counties", " candidate"]

}

);


```

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="electricity">

  <a href="electricity/" class="dataset-link float-left">

  </a>

  <a href="electricity/" class="dataset-link">
  <h3>Electricity</h3>
  </a>

  <p>
    This data set describes over 2000 U.S. electric utilities. Reported data
for 2019 includes electrical generation, distribution, revenues, and customers.
    <a href="electricity/" class="dataset-link"><em>View</em></a>
  </p>

  <small>elctricity,electric power, utilities, utility revenues</small>

  <script>
searcher.add(
  {
    'name': "electricity",
    'body': "This data set describes over 2000 U.S. electric utilities.
Reported data for 2019 includes electrical generation, distribution, revenues,
and customers.",
    'description': "The U.S. Energy Information Adminstration collects
and curates self-reported information from energy utilities about energy
production and usage in the United States. This data set contains information
from over 2,000 U.S. utilities in 2017. The information includes sources of
energy, its uses in different economic sectors, and the revenues obtained from
the sale of electrical energy.",
    'tags': ["elctricity", "electric power", " utilities", " utility
revenues"]
  }
);
</script>

```

</div>

<div class="col-4 ml-md-auto dataset-card" id="emissions">

<h3>Emissions</h3>

<p>

The Emissions Database for Atmospheric Research (EDGAR) supported by the European Union shows green house gas emissions by country. This data set contains data from 1970 through 2012.

View

</p>

<small>hospitals, health care, medical, hospital costs, hospital quality</small>

<script>

searcher.add(

{

'name': "emissions",

'body': "The Emissions Database for Atmospheric Research (EDGAR) supported by the European Union shows green house gas emissions by country. This data set contains data from 1970 through 2012.",

'description': "The data set allows consumers to directly compare across hospitals performance measure information related to heart attack, emergency department care, preventive care, stroke care, and other conditions. The data is part of an Administration-wide effort to increase the availability and accessibility of information on quality, utilization, and costs for effective, informed decision-making.",

'tags': ["hospitals", " health care", " medical", " hospital costs",
" hospital quality"]

}

);

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="energy">

<h3>Energy</h3>

<p>

United States Government reports on consumption, production, import, and export of various fuel sources.

View

</p>

<small>energy, renewable, united states, us, nuclear, fossil fuels, power</small>

<script>

searcher.add(

{

'name': "energy",

'body': "United States Government reports on consumption,

production, import, and export of various fuel sources.",

'description': "United States Government reports on consumption,

production, expenditures, and prices of various fuel sources. This data comes

from the US Energy Information Administration, which has historical data from

1960 to 2019. Information was not always available, in which case 0 was reported

instead. In some cases, very tiny values were rounded down to zero.",

'tags': ["energy", " renewable", " united states", " us", "

nuclear", " fossil fuels", " power"]

}

);

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="finance">

```

<a href="finance/" class="dataset-link float-left">



</a>

<a href="finance/" class="dataset-link">
<h3>Finance</h3>
</a>

<p>
    The Annual Survey of State Government Finances provides a comprehensive
    summary of the annual survey findings for state governments, as well as data for
    individual states.
    <a href="finance/" class="dataset-link"><em>View</em></a>
</p>

<small>finance, government, expenditure, charges, money, united states,
us</small>

<script>
searcher.add(
    {
        'name': "finance",
        'body': "The Annual Survey of State Government Finances provides a
comprehensive summary of the annual survey findings for state governments, as
well as data for individual states.",
        'description': "The Annual Survey of State Government Finances
provides a comprehensive summary of the annual survey findings for state
governments, as well as data for individual states. The data contain detail of
revenue by source, expenditure by object and function, indebtedness by term, and
assets by purpose. This data goes back for almost two decades across every date.
(1992 - 2019)",
        'tags': ["finance", " government", " expenditure", " charges", "
money", " united states", " us"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="food">

```

```


!\[Food\]\(/corgis//images/datasets/food-icon.png\)
<h3>Food</h3>
</a>

<p>
    Statistics for various food items
    <a href="food/" class="dataset-link"><em>View</em></a>
</p>

<small>food, vitamins, minerals, health, nutrition</small>

<script>
searcher.add\(
    {
        'name': "food",
        'body': "Statistics for various food items",
        'description': "The following data comes from the United States
Department of Agriculture's Food Composition Database. It contains data for
various types of food including the amounts of different vitamins and minerals
found in the foods as well as macronutrient percentages. The food covered spans
a large variety of foods from butter to Campbell's soup. Much of the
supplementary documenation for each field comes directly from that pages'
Wikipedia article.",
        'tags': \["food", " vitamins", " minerals", " health", " nutrition"\]
    }
\);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="food access">

    <a href="food\_access/" class="dataset-link float-left">

    </a>

```

```

<a href="food_access/" class="dataset-link">
<h3>Food Access</h3>
</a>

<p>
    Data about counties ability to access supermarkets, supercenters,
    grocery stores, or other sources of healthy and affordable food.
    <a href="food_access/" class="dataset-link"><em>View</em></a>
</p>

<small>counties, states, food, access, availability, supermarket,...</small>

<script>
searcher.add(
    {
        'name': "food_access",
        'body': "Data about counties ability to access supermarkets,
supercenters, grocery stores, or other sources of healthy and affordable food.",
        'description': "From the United States Department of Agriculture's
Economic Research Service, the dataset contains information about US county's
ability to access supermarkets, supercenters, grocery stores, or other sources
of healthy and affordable food. Most measures of how individuals and
neighborhoods are able to access food are based on the following indicators: -
Accessibility to sources of healthy food, as measured by distance to a store or
by the number of stores in an area. - Individual-level resources that may affect
accessibility, such as family income or vehicle availability. - Neighborhood-
level indicators of resources, such as the average income of the neighborhood
and the availability of public transportation.",
        'tags': ["counties", " states", " food", " access", " availability",
" supermarket", " rural", " urban", " population", " vehicles"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="global_development">

    <a href="global_development/" class="dataset-link float-left">

    </a>

```

```

<a href="global_development/" class="dataset-link">
<h3>Global Development</h3>
</a>

<p>
  Reports of country's development over time
  <a href="global_development/" class="dataset-link"><em>View</em></a>
</p>

<small>world, countries, development, health, population, urban,...</small>

<script>
searcher.add(
  {
    'name': "global_development",
    'body': "Reports of country's development over time",
    'description': "The following data contains records collected on
different countries and geographic locations from 1980 - 2013 from the World
Bank. Included is different data about urban development, agriculture and rural
development, health, and infrastructure.",
    'tags': ["world", " countries", " development", " health", "
population", " urban", " rural", " agriculture", " infrastructure"]
  }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="graduates">

  <a href="graduates/" class="dataset-link float-left">

  </a>

  <a href="graduates/" class="dataset-link">
  <h3>Graduates</h3>
  </a>

  <p>
    This library holds data about employment of new graduates by major.
    <a href="graduates/" class="dataset-link"><em>View</em></a>
  
```

```

</p>

<small>graduation, grads, majors, education, college, university,...</small>

<script>
searcher.add(
    {
        'name': "graduates",
        'body': "This library holds data about employment of new graduates
by major.",
        'description': "The data in this library comes from the National
Survey of Recent College Graduates. Included is information about employment
numbers, major information, and the earnings of different majors. Many majors
were not available before 2010, so their values have been recorded as 0 (note
that this may affect the averages shown in the bar charts).",
        'tags': ["graduation", " grads", " majors", " education", "
college", " university", " higher education", " jobs", " careers"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="health">

    <a href="health/" class="dataset-link float-left">

    </a>

    <a href="health/" class="dataset-link">
    <h3>Health</h3>
    </a>

    <p>
        This library holds data about reports of diseases in America.
        <a href="health/" class="dataset-link"><em>View</em></a>
    </p>

    <small>health, diseases, infection</small>

    <script>
searcher.add(

```



```

        {
            'name': "health",
            'body': "This library holds data about reports of diseases in
America.",
            'description': "This library holds data about reports of diseases in
America.",
            'tags': ["health", " diseases", " infection"]
        }
    );
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="hospitals">

```

```

    <a href="hospitals/" class="dataset-link float-left">

```

```

```

```

    </a>

```

```

    <a href="hospitals/" class="dataset-link">

```

```

        <h3>Hospitals</h3>

```

```

    </a>

```

```

    <p>

```

```

        To help consumers make informed decisions about health care, the Centers
for Medicare & Medicaid Services (CMS) collects data about the cost and
quality of care at over 4,000 Medicare-qualified hospitals.

```

```

        <a href="hospitals/" class="dataset-link"><em>View</em></a>

```

```

    </p>

```

```

    <small>hospitals, health care, medical, hospital costs, hospital
quality</small>

```

```

    <script>

```

```

        searcher.add(

```

```

            {

```

```

                'name': "hospitals",

```

```

                'body': "To help consumers make informed decisions about health
care, the Centers for Medicare & Medicaid Services (CMS) collects data about the
cost and quality of care at over 4,000 Medicare-qualified hospitals.",

```

```

                'description': "The data set allows consumers to directly compare
across hospitals performance measure information related to heart attack,

```

emergency department care, preventive care, stroke care, and other conditions. The data is part of an Administration-wide effort to increase the availability and accessibility of information on quality, utilization, and costs for effective, informed decision-making.",

```
        'tags': ["hospitals", " health care", " medical", " hospital costs",  
" hospital quality"]  
    }  
);  
</script>
```

</div>

```
<div class="col-4 ml-md-auto dataset-card" id="hydropower">
```

```
    <a href="hydropower/" class="dataset-link float-left">
```

```
        
```

```
    </a>
```

```
    <a href="hydropower/" class="dataset-link">
```

```
        <h3>Hydropower</h3>
```

```
    </a>
```

```
    <p>
```

Records about dams in the United States such as location, dimensions,
and project information

```
        <a href="hydropower/" class="dataset-link"><em>View</em></a>
```

```
    </p>
```

```
    <small>hydropower, dams, lakes, rivers, waterways, irrigation,...</small>
```

```
    <script>
```

```
        searcher.add(  
            {
```

```
                'name': "hydropower",
```

```
                'body': "Records about dams in the United States such as location,  
dimensions, and project information",
```

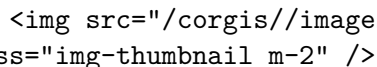
```
                'description': "Recent Federal policies and legislation focus on  
moving the nation towards a cleaner energy economy that includes developing  
environmentally appropriate renewable energy projects involving solar, wind and  
waves, geothermal, biofuels, and hydropower. The 2010 Federal Memorandum of  
Understanding for Hydropower and the Energy Policy Act of 2005 direct the U.S.  
Department of Reclamation to evaluate development of new hydropower projects at
```

Federally-owned facilities and upgrade or rehabilitate existing hydropower generation facilities, as a contribution to the nation's clean energy goals. The following data describes the facilities currently being operated that contribute to nationwide renewable energy strategies.",

```
    'tags': ["hydropower", " dams", " lakes", " rivers", " waterways", "
irrigation", " human consumption", " industrial use", " aquaculture", "
navigability", " energy", " electricity"]
  }
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="ingredients">

Ingredients" class="img-thumbnail m-2" />

Ingredients

<p>

Statistics for various food ingredients

[View](ingredients/)

</p>

ingredients, vitamins, minerals, health, nutrition

<script>

searcher.add(

{

'name': "ingredients",

'body': "Statistics for various food ingredients",

'description': "The following data comes from the United States Department of Agriculture's Food Composition Database. It contains data for various types of food ingredients including the amounts of different vitamins and minerals found in the foods as well as macronutrient percentages. The food covered spans a large variety of foods from butter to Campbell's soup. Much of the supplementary documentation for each field comes directly from that pages' Wikipedia article.",

```

        'tags': ["ingredients", " vitamins", " minerals", " health", "
nutrition"]
    }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="injuries">

```

```

    <a href="injuries/" class="dataset-link float-left">

```

```

```

```

    </a>

```

```

    <a href="injuries/" class="dataset-link">
    <h3>Injuries</h3>
    </a>

```

```

    <p>
        A sampling of work-related injury and illness data from the US.
        <a href="injuries/" class="dataset-link"><em>View</em></a>
    </p>

```

```

    <small>osha, united states, us, usa, government, health, safety,...</small>

```

```

    <script>
    searcher.add(
        {
            'name': "injuries",
            'body': "A sampling of work-related injury and illness data from the
US.",
            'description': "The Occupational Safety and Health Administration
(OSHA) collected work-related injury and illness data from employers within
specific industry and employment size specifications from 2002 through 2011.
This data collection is called the OSHA Data Initiative or ODI. The data
provided is used by OSHA to calculate establishment specific injury and illness
incidence rates. This searchable database contains a table with the name,
address, industry, and associated Total Case Rate (TCR), Days Away, Restricted,
and Transfer (DART) case rate, and the Days Away From Work (DAFWII) case rate
for the establishments that provided OSHA with valid data for calendar years
2002 through 2011. This data has been sampled down from its original size to 4%.
In addition, the original dataset only has data from a small portion of all

```

private sector establishments in the United States (80,000 out of 7.5 million total establishments). Therefore, these data are not representative of all businesses and general conclusions pertaining to all US business should not be overdrawn. Data quality: While OSHA takes multiple steps to ensure the data collected is accurate, problems and errors invariably exist for a small percentage of establishments. OSHA does not believe the data for the establishments with the highest rates on this file are accurate in absolute terms. Efforts were made during the collection cycle to correct submission errors, however some remain unresolved. It would be a mistake to say establishments with the highest rates on this file are the 'most dangerous' or 'worst' establishments in the Nation. Rate Calculation: An incidence rate of injuries and illnesses is computed from the following formula: (Number of injuries and illnesses X 200,000) / Employee hours worked = Incidence rate. The Total Case Rate includes all cases recorded on the OSHA Form 300 (Column G + Column H + Column I + Column J). The Days Away/Restricted/Transfer includes cases recorded in Column H + Column I. The Days Away includes cases recorded in Column H. For further information on injury and illness incidence rates, please visit the Bureau of Labor Statistics' webpage at <http://www.bls.gov/iif/osheval.htm> State Participation: Not all state plan states participate in the ODI. The following states did not participate in the 2010 ODI (collection of CY 2009 data), establishment data is not available for these states: Alaska; Oregon; Puerto Rico; South Carolina; Washington; Wyoming.",

```

    'tags': ["osha", " united states", " us", " usa", " government", "
health", " safety", " injury", " injuries", " hurt", " pain", " job", " work", "
industry", " career", " occupation", " safety", " hazard", " health"]
}
);
</script>

```

</div>

<div class="col-4 ml-md-auto dataset-card" id="labor">

<h3>Labor</h3>

<p>

Employment status of the civilian population by race, sex, and age.
[View](labor/)

</p>

<small>labor, race, age, sex, gender, america, usa, census,...</small>

<script>

searcher.add(

{

'name': "labor",

'body': "Employment status of the civilian population by race, sex, and age.",

'description': "The Current Population Survey (CPS) is a monthly survey of households conducted by the Bureau of Census for the Bureau of Labor Statistics. It provides a comprehensive body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics.",

'tags': ["labor", " race", " age", " sex", " gender", " america", " usa", " census", " employed", " unemployed", " employability", " job", " work", " civilian", " black", " white", " asian", " government"]

}

);

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="medal of honor">

![Medal Of Honor](/corgis//images/datasets/medal-of-honor-icon.png)

<h3>Medal Of Honor</h3>

<p>

This dataset has records for the awarding of the United States Medal of Honor, one of the military's highest honors.

[View](medal_of_honor/)

</p>

```
<small>us, usa, united states, military, combat, government,  
army,...</small>
```

```
<script>  
searcher.add(  
  {  
    'name': "medal_of_honor",  
    'body': "This dataset has records for the awarding of the United  
States Medal of Honor, one of the military's highest honors.",  
    'description': "This dataset has records for the awarding of the  
United States Medal of Honor. The Medal of Honor is the United States of  
America's highest military honor, awarded for personal acts of valor above and  
beyond the call of duty. The medal is awarded by the President of the United  
States in the name of the U.S. Congress to U.S. military personnel only. There  
are three versions of the medal, one for the Army, one for the Navy, and one for  
the Air Force.[5] Personnel of the Marine Corps and Coast Guard receive the Navy  
version. The dataset was collected from the official military site, and includes  
records about how the medal was awarded and characteristics of the recipient.  
Unfortunately, because of the nature of century-old record keeping, many of the  
records are incomplete. While a very interesting dataset, it does have some  
missing data.",  
    'tags': ["us", " usa", " united states", " military", " combat", " "  
government", " army", " airforce", " marines", " navy", " fighting", " war", "  
award", " medal", " honor", " glory", " death"]  
  }  
);  
</script>
```

```
</div>
```

```
<div class="col-4 ml-md-auto dataset-card" id="music">
```

```
<a href="music/" class="dataset-link float-left">
```

```

```

```
</a>
```

```
<a href="music/" class="dataset-link">
```

```
<h3>Music</h3>
```

```
</a>
```

```
<p>
```

The Music library is a compilation of over one million contemporary

songs and information about their audio features and metadata.

[View](music/)

</p>

<small>music, songs, artists, creativity, media</small>

<script>

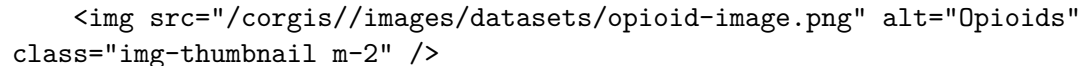
```
searcher.add(
  {
    'name': "music",
    'body': "The Music library is a compilation of over one million
contemporary songs and information about their audio features and metadata.",
    'description': "This library comes from the Million Song Dataset,
which used a company called the Echo Nest to derive data points about one
million popular contemporary songs. The Million Song Dataset is a collaboration
between the Echo Nest and LabROSA, a laboratory working towards intelligent
machine listening. The project was also funded in part by the National Science
Foundation of America (NSF) to provide a large data set to evaluate research
related to algorithms on a commercial size while promoting further research into
the Music Information Retrieval field. The data contains standard information
about the songs such as artist name, title, and year released. Additionally, the
data contains more advanced information; for example, the length of the song,
how many musical bars long the song is, and how long the fade in to the song
was.",
    'tags': ["music", " songs", " artists", " creativity", " media"]
  }
);
```

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="opioids">

[](opioids/)



[](opioids/)

<h3>Opioids</h3>

<p>

56

The data set shows the number and rates of deaths due to opioid overdose.

[View](opioids/)

</p>

<small>opioids, drug deaths, heroin, cocaine, drug overdose</small>

<script>

searcher.add(

{

'name': "opioids",

'body': "The data set shows the number and rates of deaths due to opioid overdose.",

'description': "All opioids are chemically related and interact with opioid receptors on nerve cells in the body and brain. Opioid pain relievers can be misused (taken in a different way or in a larger quantity than prescribed, or taken without a doctor's prescription). Regular use - even as prescribed by a doctor - can lead to dependence and, when misused, opioid pain relievers can lead to addiction, overdose incidents, and deaths. The National Institute on Drug Abuse collects and analyzes data about deaths from opioid abuse. This data set reports on data from 1999-2019.",

'tags': ["opioids", " drug deaths", " heroin", " cocaine", " drug overdose"]

}

);

</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="police shootings">

[](police_shootings/)

[](police_shootings/)

<h3>Police Shootings</h3>

<p>

The Washington Post is compiling a database of every fatal shooting in the United States by a police officer in the line of duty since Jan. 1, 2015 by

culling local news reports, law enforcement websites and social media and by monitoring independent databases. The Post conducted additional reporting in many cases.

[View
</p>](police_shootings/)

<small>violence, crime, violent, police, shootings, race, location,
time</small>

```
<script>
searcher.add(
  {
    'name': "police_shootings",
    'body': "The Washington Post is compiling a database of every fatal
shooting in the United States by a police officer in the line of duty since Jan.
1, 2015 by culling local news reports, law enforcement websites and social media
and by monitoring independent databases. The Post conducted additional reporting
in many cases.",
    'description': "https://www.washingtonpost.com/national/how-the-
washington-post-is-examining-police-shootings-in-the-united-
states/2016/07/07/d9c52238-43ad-11e6-8856-f26de2537a9d_story.html",
    'tags': ["violence", " crime", " violent", " police", " shootings",
" race", " location", " time"]
  }
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="publishers">

![Publishers](/corgis//images/datasets/publishers-icon.png)

<h3>Publishers</h3>

<p>
E-book sales on Amazon, including daily and total earnings for 54,000 titles.

```

    <a href="publishers/" class="dataset-link"><em>View</em></a>
</p>

<small>publishers, amazon, books, sales, genres, literature, english</small>

<script>
searcher.add(
    {
        'name': "publishers",
        'body': "E-book sales on Amazon, including daily and total earnings
for 54,000 titles.",
        'description': "From a newspaper article about analyzing amazon
e-book sales by genre and publisher. Unfortunately, they do not have information
on the book's title or author. This collection includes 54,000 titles spanning
across several genres and types of publishing companies, practically every book
on every Amazon bestseller list. Along with publisher information, it also
includes the book's overall Amazon Kindle store sales ranking. This ranking is
used to sort the books. Keep in mind that this data is NOT time-oriented; it is
a collection of a bunch of different books, not a book over time.",
        'tags': ["publishers", " amazon", " books", " sales", " genres", "
literature", " english"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="real estate">

    <a href="real_estate/" class="dataset-link float-left">

    </a>

    <a href="real_estate/" class="dataset-link">
<h3>Real Estate</h3>
</a>

    <p>
        Real estate information in the United States, including inventory,
building, and customer data.
        <a href="real_estate/" class="dataset-link"><em>View</em></a>

```

```

</p>

<small>real, estate, buildings, government, building</small>

<script>
searcher.add(
    {
        'name': "real_estate",
        'body': "Real estate information in the United States, including
inventory, building, and customer data.",
        'description': "Real Estate Across the United States (REXUS) is the
primary tool used by PBS to track and manage the government's real property
assets and to store inventory data, building data, customer data, and lease
information. STAR manages aspects of real property space management, including
identification of all building space and daily management of 22,000 assignments
for all property to its client Federal agencies. This data set contains PBS
building inventory that consists of both owned and leased buildings with active
and excess status.",
        'tags': ["real", " estate", " buildings", " government", "
building"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="retail services">

    <a href="retail_services/" class="dataset-link float-left">

    </a>

    <a href="retail_services/" class="dataset-link">
    <h3>Retail Services</h3>
    </a>

    <p>
        Monthly sales, inventories, and inventory/sales ratios for retail trade
and food services.
        <a href="retail_services/" class="dataset-link"><em>View</em></a>
    </p>

```

<small>retail, services, government, united states, usa, us,
trade,...</small>

```
<script>
searcher.add(
  {
    'name': "retail_services",
    'body': "Monthly sales, inventories, and inventory/sales ratios for
retail trade and food services.",
    'description': "The Advance Monthly and Monthly Retail Trade Surveys
(MARTS and MRTS), the Annual Retail Trade Survey (ARTS), and the Quarterly
E-Commerce Report work together to produce the most comprehensive data available
on retail economic activity in the United States. More detailed descriptions of
these programs can be found by choosing one of the links to the left. Regular
quality control and verification takes place between MARTS, MRTS, and ARTS
annually and between these programs and the Economic Census of Retail Trade
every five years. Each year when annual data become available, we compare and
resolve differences between the data collected on the monthly and annual
surveys. We refer to this process as the monthly-to-annual reconciliation. At
the same time, we benchmark the monthly estimates using results of the annual
survey. ARTS estimates are then benchmarked to data maintained by the Economic
Census of Retail Trade. This process of benchmarking retail data over all four
programs ensures consistency in our estimates. The Business Expenses Supplement
is an addition to the 2007 Annual Retail Trade Survey. Its purpose is to compile
statistics on detailed business operating expenses. The United States Code,
Title 13, authorizes this program as part of the Economic Census. This coverage
was previously part of the predecessor Business Expenses Survey. Detailed
expenses are collected every five years.",
    'tags': ["retail", " services", " government", " united states", "
usa", " us", " trade", " food", " services"]
  }
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="school scores">


```

<a href="school_scores/" class="dataset-link">
<h3>School Scores</h3>
</a>

<p>
    SAT scores for students across the United States, organized by year and
state.
    <a href="school_scores/" class="dataset-link"><em>View</em></a>
</p>

<small>school, education, K-12, kindergarten, learning, teaching,
sat</small>

<script>
searcher.add(
    {
        'name': "school_scores",
        'body': "SAT scores for students across the United States, organized
by year and state.",
        'description': "SAT Scores from across the country",
        'tags': ["school", " education", " K-12", " kindergarten", "
learning", " teaching", " sat"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="skyscrapers">

    <a href="skyscrapers/" class="dataset-link float-left">

    </a>

    <a href="skyscrapers/" class="dataset-link">
<h3>Skyscrapers</h3>
</a>

<p>
    Information about tall buildings constructed all over the world.
    <a href="skyscrapers/" class="dataset-link"><em>View</em></a>
</p>

```

```

<small>buildings, construction, building, skyscrapers,...</small>

<script>
searcher.add(
  {
    'name': "skyscrapers",
    'body': "Information about tall buildings constructed all over the
world.",
    'description': "Collected statistics about the tallest buildings all
over the world, including the materials used to develop them, the purpose of the
buildings, and more.",
    'tags': ["buildings", " construction", " building", " skyscrapers",
" international", " tall", " height", " tower"]
  }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="slavery">

  <a href="slavery/" class="dataset-link float-left">

  </a>

  <a href="slavery/" class="dataset-link">
<h3>Slavery</h3>
</a>

  <p>
    This dataset has financial records of New Orleans slave sales,
1856-1861. There are almost 16,000 sales recorded in this dataset.
    <a href="slavery/" class="dataset-link"><em>View</em></a>
  </p>

  <small>slavery, slave, slaves, buyer, seller, origin, history,
economics</small>

  <script>
searcher.add(
  {

```

```

        'name': "slavery",
        'body': "This dataset has financial records of New Orleans slave
sales, 1856-1861. There are almost 16,000 sales recorded in this dataset.",
        'description': "This dataset has financial records of New Orleans
slave sales, 1856-1861. There are almost 16,000 sales recorded in this
dataset.",
        'tags': ["slavery", " slave", " slaves", " buyer", " seller", "
origin", " history", " economics"]
    }
);
</script>

```

</div>

```
<div class="col-4 ml-md-auto dataset-card" id="state crime">
```

```
    <a href="state_crime/" class="dataset-link float-left">
```

```

```

```
    </a>
```

```
    <a href="state_crime/" class="dataset-link">
```

```
        <h3>State Crime</h3>
```

```
    </a>
```

```
    <p>
```

```
        Records about the crime rates and totals for US states over time.
```

```
        <a href="state_crime/" class="dataset-link"><em>View</em></a>
```

```
    </p>
```

```

    <small>crime, burglary, larceny, motor, property, violent,
assault,...</small>

```

```
    <script>
```

```
    searcher.add(
```

```
        {
```

```
            'name': "state_crime",
```

```

            'body': "Records about the crime rates and totals for US states over
time.",

```

```

            'description': "From the Unified Crime Reporting Statistics and
under the collaboration of the U.S. Department of Justice and the Federal Bureau
of Investigation information crime statistics are available for public review.
The following data set has information on the crime rates and totals for states

```


across the United States for a wide range of years. The crime reports are divided into two main categories: property and violent crime. Property crime refers to burglary, larceny, and motor related crime while violent crime refers to assault, murder, rape, and robbery. These reports go from 1960 to 2019.",

```
    'tags': ["crime", " burglary", " larceny", " motor", " property", "
violent", " assault", " murder", " rape", " robbery"]
  }
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="state demographics">

 <h3>State Demographics</h3>

 <p>

 Demographic information for the states in the United States.

 View

 </p>

 <small>demographics, population, age, ethnicity, race, language,...</small>

 <script>

 searcher.add(

 {

 'name': "state_demographics",

 'body': "Demographic information for the states in the United

States.",

 'description': "The following data set is summarized information

obtained about states in the United States from 2015 through 2019 through the United States Census Bureau. However, this data is not available for these years; just the summarized data as of 2019. Information described in the data includes the age distributions, the education levels, employment statistics, ethnicity percents, household information, income, and other miscellaneous statistics. (Values are denoted as -1, if the data is not available)",

```

        'tags': ["demographics", " population", " age", " ethnicity", "
race", " language", " education", " foreign", " households", " income", "
employment", " unemployment", " sales", " land area", " states"]
    }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="state fragility">

```

```

    <a href="state_fragility/" class="dataset-link float-left">

```

```

```

```

    </a>

```

```

    <a href="state_fragility/" class="dataset-link">

```

```

        <h3>State Fragility</h3>

```

```

    </a>

```

```

    <p>

```

```

        The State Fragility Index scores for countries around the world from
1995 to 2018

```

```

        <a href="state_fragility/" class="dataset-link"><em>View</em></a>

```

```

    </p>

```

```

    <small>world, countries, security, politics, economy, society,...</small>

```

```

    <script>

```

```

    searcher.add(
        {
            'name': "state_fragility",
            'body': "The State Fragility Index scores for countries around the
world from 1995 to 2018",
            'description': "The Center for Systemic Peace was founded in 1997 to
engage in global systems analysis to minimize the effects of political violence
in the world as a whole. The following data set shows the state fragility for
countries with a population greater than 500,000 in 2013. The State Fragility
Index scores countries on two main categories, effectiveness and legitimacy.
These are then broken down into four dimensions: Security, Political, Economic,
and Social. The State Fragility Index score is the sum of the individual
country's effectiveness score and their legitimacy score. Each of these scores
are the summation of the four dimensions within the category.",

```

```

        'tags': ["world", " countries", " security", " politics", "
economy", " society", " effectiveness", " legitimacy"]
    }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="suicide attacks">

```

```

    <a href="suicide_attacks/" class="dataset-link float-left">

```

```

```

```

    </a>

```

```

    <a href="suicide_attacks/" class="dataset-link">
    <h3>Suicide Attacks</h3>
    </a>

```

```

    <p>
        This dataset has about suicide attacks all over the world for the past
three decades.

```

```

        <a href="suicide_attacks/" class="dataset-link"><em>View</em></a>
    </p>

```

```

    <small>global, attack, damage, kill, hurt, destroy, pain,...</small>

```

```

    <script>
    searcher.add(
        {
            'name': "suicide_attacks",
            'body': "This dataset has about suicide attacks all over the world
for the past three decades.",
            'description': "The Chicago Project on Security and Terrorism
(CPOST) maintains a searchable database on all suicide attacks from 1982 through
October 2020. The database includes information about the location of attacks,
the target type, the weapon used, and systematic information on the demographic
and general biographical characteristics of suicide attackers. The database
expands the breadth of the data available in English using native language
sources (e.g., Arabic, Hebrew, Russian, Tamil) that are likely to have the most
extensive relevant information. The current CPOST-SAD release contains the
universe of suicide attacks from 1982 through September 2015, a total of 4,814
attacks in over 40 countries. The CPOST Suicide Attack Database (CPOST-SAD)

```

contains data on the universe of suicide attacks. For each attack, the database includes information about the geographic location, target classification, and weapon(s) used, as well as systematic information on the demographic and general biographical characteristics of suicide attackers. More information about the coding and methodology is available on the CPOST website below. Note that multiple attackers or targets are collapsed into a single record for simplicity's sake.",

```
    'tags': ["global", " attack", " damage", " kill", " hurt", "
destroy", " pain", " suffering", " terror", " terrorism", " religion", "
warfare", " war", " car", " bomb", " gun", " shooting"]
  }
```

```
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="supreme court">

 <h3>Supreme Court</h3>

 <p>

 This dataset contains information about supreme court cases for the several decades.

 View

 </p>

 <small>us, usa, united states, government, law, legal, court, case,...</small>

 <script>

```
searcher.add(
  {
    'name': "supreme_court",
    'body': "This dataset contains information about supreme court cases
for the several decades.",
    'description': "The U.S. Supreme Court Database traces its history
```

back about two decades ago, when Harold J. Spaeth asked the National Science Foundation to fund a database that would be so rich in content that multiple users - even those with vastly distinct projects and purposes in mind - could draw on it. Professor Spaeth's goal was at once refreshingly simple and extremely ambitious: to produce a database that would include and classify every single vote by a Supreme Court justice in all argued cases over a five-decade period. After securing the funding, Spaeth collected and coded the data, performed reliability checks, and eventually amassed the Database. In the late 1980s, he made it (and the documentation necessary to use it) publicly available. Since then, Professor Spaeth has not only updated it each term; he has also continued to perform reliability analyses, thereby ensuring its integrity with each release, and added new variables. Today's version of the Database houses 247 pieces of information for each case, roughly broken down into six categories: (1) identification variables (e.g., citations and docket numbers); (2) background variables (e.g., how the Court took jurisdiction, origin and source of the case, the reason the Court agreed to decide it); (3) chronological variables (e.g., the date of decision, term of Court, natural court); (4) substantive variables (e.g., legal provisions, issues, direction of decision); (5) outcome variables (e.g., disposition of the case, winning party, formal alteration of precedent, declaration of unconstitutionality); and (6) voting and opinion variables (e.g., how the individual justices voted, their opinions and interagreements). Case Centered data provides case level information; i.e., each row in the database corresponds to a dispute. These data do not contain specific justice vote information. A tremendously helpful resource for learning more about the dataset is the codebook:

<http://supremecourtdatabase.org/documentation.php?s=1.>

```
'tags': ["us", " usa", " united states", " government", " law", "
legal", " court", " case", " supreme", " judge", " barrister", " lawyer", "
federal", " national"]
}
);
</script>
```

</div>

<div class="col-4 ml-md-auto dataset-card" id="tate">


```

<h3>Tate</h3>
</a>

<p>
    This dataset is about the Tate art collection, with metadata about
    paintings, drawings, sculptures, and more.
    <a href="tate/" class="dataset-link"><em>View</em></a>
</p>

<small>artwork, art, sculpture, tate</small>

<script>
searcher.add(
    {
        'name': "tate",
        'body': "This dataset is about the Tate art collection, with
        metadata about paintings, drawings, sculptures, and more.",
        'description': "Tate holds the national collection of British art
        from 1500 to the present day and international modern and contemporary art. The
        collection embraces all media, from painting, drawing, sculpture and prints to
        photography, video and film, installation and performance. This dataset is not
        ordered in any particular way. This dataset has metadata about many pieces from
        their collection.",
        'tags': ["artwork", " art", " sculpture", " tate"]
    }
);
</script>

</div>

<div class="col-4 ml-md-auto dataset-card" id="video games">

    <a href="video_games/" class="dataset-link float-left">

    </a>

    <a href="video_games/" class="dataset-link">
    <h3>Video Games</h3>
    </a>

    <p>
        This library has data about video games, including their play time and

```

```

sales.
    <a href="video_games/" class="dataset-link"><em>View</em></a>
</p>

<small>video games, games, publishers, console, gaming</small>

<script>
searcher.add(
    {
        'name': "video_games",
        'body': "This library has data about video games, including their
play time and sales.",
        'description': "This dataset requires special permission to use.
Originally collected by Dr. Joe Cox, this dataset has information about the
sales and playtime of over a thousand video games released between 2004 and
2010. The playtime information was collected from crowd-sourced data on \"How
Long to Beat\". Some more information can be found here.",
        'tags': ["video games", " games", " publishers", " console", "
gaming"]
    }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="weather">

    <a href="weather/" class="dataset-link float-left">

    </a>

    <a href="weather/" class="dataset-link">
<h3>Weather</h3>
</a>

    <p>
        Weekly weather records in 2016 across the U.S.
        <a href="weather/" class="dataset-link"><em>View</em></a>
    </p>

    <small>weather, rain, snow, sleet, fog, temperature, wind,
climate,...</small>

```

```

<script>
searcher.add(
  {
    'name': "weather",
    'body': "Weekly weather records in 2016 across the U.S.",
    'description': "Under the National Oceanic and Atmospheric
Administration, the National Weather Service provides daily weather reports for
cities across the county. This is done through the use of 122 different Weather
Forecast Offices throughout the country. These WFOs are responsible for the daily
weather reports for serveral cities throughout their region of coverage. This
data set takes the information from these WFO reports for cities across the
country and summarizes it at the weekly level for all of 2016.",
    'tags': ["weather", " rain", " snow", " sleet", " fog", "
temperature", " wind", " climate", " environment", " geology"]
  }
);
</script>

```

```

</div>

```

```

<div class="col-4 ml-md-auto dataset-card" id="wind turbines">

```

```

  <a href="wind_turbines/" class="dataset-link float-left">

```

```

```

```

  </a>

```

```

  <a href="wind_turbines/" class="dataset-link">

```

```

    <h3>Wind Turbines</h3>

```

```

  </a>

```

```

  <p>

```

This dataset provides locations and technical specifications of wind turbines in the United States, almost all of which are utility-scale. Utility-scale turbines are ones that generate power and feed it into the grid, supplying a utility with energy. They are usually much larger than turbines that would feed a homeowner or business.

```

    <a href="wind_turbines/" class="dataset-link"><em>View</em></a>

```

```

  </p>

```

```

  <small>energy, wind, power, air, turbine, utility, location, time</small>

```



```

<script>
searcher.add(
  {
    'name': "wind_turbines",
    'body': "This dataset provides locations and technical
specifications of wind turbines in the United States, almost all of which are
utility-scale. Utility-scale turbines are ones that generate power and feed it
into the grid, supplying a utility with energy. They are usually much larger
than turbines that would feed a homeowner or business.",
    'description': "The Open PV Project is a collaborative effort
between government, industry, and the public that continues to compile a
database of available public data for photovoltaic (PV) installation data for
the United States. Data for the project are voluntarily contributed from a
variety of sources including solar incentive programs, utilities, installers,
and the general public. This database serves as a web-based resource for users
to easily explore and understand the current and past trends of the US PV
industry. The data collected is actively maintained by the contributors and are
always changing to provide an evolving, up-to-date snapshot of the US solar
power market.",
    'tags': ["energy", " wind", " power", " air", " turbine", "
utility", " location", " time"]
  }
);
</script>

```

</div>

</div>

<hr />

</div>

```

<footer id="footer" class="site-footer">
<div class="copyright">

```

<p>© 2022 CORGIS Datasets Project. Project by Austin Cory Bart, Dennis Kafura, Clifford A. Shaffer, Javier Tibau, Luke Gusukuma, Eli Tilevich.</p>

```

</div>
</footer>

```

```

<script
src="https://use.fontawesome.com/releases/v5.0.12/js/all.js"></script>

```

```

<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/js/bootstrap.bundle.min.js"></script>
<script type="text/javascript">
  $(document).ready(function() {
    $("#search-bar").on('input propertychange', function() {
      let needle = $(this).val();
      if (needle === "") {
        $(".dataset-card").show();
      } else {
        let results = searcher.search(needle);
        $(".dataset-card").hide();
        results.forEach(function (data) {
          $("#"+data.ref).show();
        });
      }
    });
  });
</script>

</body>

</html>

```

```
[5]: from bs4 import BeautifulSoup
```

```

[6]: # Save all the html in a string variable
html_string = response.text

# Use BeautifulSoup to create a new object that will allow you to search for
↳HTML tags
document = BeautifulSoup(html_string, "html.parser")

# This "document" variable is an object that has a "find" method
h3 = document.find_all('h3')
for corgis in h3:
    print (corgis.string)

```

Aids
 Airlines
 Billionaires
 Broadway
 Business Dynamics
 Cancer
 Cars
 Classics
 Construction Permits
 Construction Spending
 County Demographics

Covid
Drugs
Earthquakes
Election
Electricity
Emissions
Energy
Finance
Food
Food Access
Global Development
Graduates
Health
Hospitals
Hydropower
Ingredients
Injuries
Labor
Medal Of Honor
Music
Opioids
Police Shootings
Publishers
Real Estate
Retail Services
School Scores
Skyscrapers
Slavery
State Crime
State Demographics
State Fragility
Suicide Attacks
Supreme Court
Tate
Video Games
Weather
Wind Turbines

Write a function that takes an element from the list of CORGIS datasets, searches the respective CORGIS page for the CSV download link, and returns a Pandas dataframe.

- If I hypothetically defined a function with:
 - `def pdcorgis(a):`
- and wanted to call this pdcorgis function with Python code like:
 - `df = pdcorgis("Classics")`
- then an example flow of python commands inside the function could be:
 - Convert "Classics" to lowercase (using the `lower()` method)
 - Use requests with "`https://corgis-edu.github.io/corgis/csv/`" plus "classics" plus "/" to get the website's HTML

- Use BeautifulSoup to search for all the HTML link elements on that page (links have the tag “a”)
- For any link element “i”, check if i.has_attr(‘download’) is true, and then get part of the download link for the csv file via i[‘href’]
- The complete download link for the file will be “https://corgis-edu.github.io/corgis/csv/” plus “classics” plus “/” plus the result you found for i[‘href’]
- Use return corgisdf at the end of your function, where corgisdf is assigned to be the value returned by pd.read_csv() with the download link you have found

```
[7]: import pandas as pd
import matplotlib.pyplot as plt
```

```
[8]: def pdcorgis(element):
    element = element.lower()
    url = "https://corgis-edu.github.io/corgis/csv/" + element + "/"
    response = requests.get(url)
    html_string = response.text
    document = BeautifulSoup(html_string, "html.parser")
    corgis = document.find_all('a')
    for i in corgis:
        if i.has_attr('download'):
            corgisdf = pd.read_csv(url + i.attrs['href'])
    return corgisdf
```

```
[9]: #testing the function
pdcorgis("Classics")
```

```
[9]: bibliography.congress classifications bibliography.languages \
0                                PR                                en
1                                PS                                en
2                                PZ,PR                            en
3                                PR                                en
4                                PT                                en
...                               ...                               ...
1001                             NaN                             en
1002                             NaN                             en
1003                             NaN                             en
1004                             NaN                             en
1005                             NaN                             en
```

```
                                bibliography.subjects \
0  Sisters -- Fiction,Courtship -- Fiction,Social...
1  Mentally ill women -- Fiction,Feminist fiction...
2                                Fantasy
3  Monsters -- Fiction;Frankenstein's monster (Fi...
4  Psychological fiction,Metamorphosis -- Fiction
```

...	...
1001	NaN
1002	NaN
1003	NaN
1004	NaN
1005	NaN

	bibliography.title	bibliography.type	\
0	Pride and Prejudice	Text	
1	The Yellow Wallpaper	Text	
2	Alice's Adventures in Wonderland	Text	
3	Frankenstein; Or, The Modern Prometheus	Text	
4	Metamorphosis	Text	
...	
1001	The Reluctant Heroes	Text	
1002	Film Truth; September, 1920	Text	
1003	Othmar	Text	
1004	Church History (Volumes 1-3)	Text	
1005	The Uncensored Letters of a Canteen Girl	Text	

	metadata.downloads	metadata.id	metadata.rank	\
0	36576	1342	1	
1	26363	1952	2	
2	18882	11	3	
3	17128	84	4	
4	15683	5200	5	
...	
1001	0	51483	1002	
1002	0	51484	1003	
1003	0	51487	1004	
1004	0	51491	1005	
1005	0	51495	1006	

	metadata.url	bibliography.author.birth	...	\
0	https://www.gutenberg.org/ebooks/1342	1775	...	
1	https://www.gutenberg.org/ebooks/1952	1860	...	
2	https://www.gutenberg.org/ebooks/11	1832	...	
3	https://www.gutenberg.org/ebooks/84	1797	...	
4	https://www.gutenberg.org/ebooks/5200	1883	...	
...	
1001	https://www.gutenberg.org/ebooks/51483	1926	...	
1002	https://www.gutenberg.org/ebooks/51484	0	...	
1003	https://www.gutenberg.org/ebooks/51487	1839	...	
1004	https://www.gutenberg.org/ebooks/51491	1809	...	
1005	https://www.gutenberg.org/ebooks/51495	0	...	

	metrics.sentiments.polarity	metrics.sentiments.subjectivity	\
--	-----------------------------	---------------------------------	---

0	0.136713	0.522239
1	0.054174	0.534787
2	0.041079	0.497276
3	0.100902	0.539516
4	0.041997	0.479019
...
1001	0.076203	0.451757
1002	0.143879	0.485782
1003	0.099478	0.548289
1004	0.112242	0.440014
1005	0.059623	0.435443

	metrics.statistics.average letter per word \
0	4.83
1	4.41
2	4.65
3	4.77
4	4.56
...	...
1001	4.68
1002	5.02
1003	4.66
1004	5.31
1005	4.73

	metrics.statistics.average sentence length \
0	18.0
1	15.0
2	17.0
3	23.0
4	27.0
...	...
1001	11.0
1002	17.0
1003	21.0
1004	17.0
1005	17.0

	metrics.statistics.average sentence per word \
0	0.05
1	0.06
2	0.06
3	0.04
4	0.04
...	...
1001	0.09
1002	0.06

1003	0.05
1004	0.06
1005	0.06

	metrics.statistics.characters	metrics.statistics.polysyllables \
0	586794	4603
1	26769	102
2	122719	339
3	357604	2604
4	100372	397
...
1001	32354	112
1002	45970	312
1003	890918	4808
1004	3883244	46730
1005	451687	2595

	metrics.statistics.sentences	metrics.statistics.syllables \
0	6511	170648.1
1	385	7686.9
2	1501	33810.3
3	3239	106802.1
4	800	28752.3
...
1001	618	8824.5
1002	536	13080.6
1003	8921	258091.2
1004	41950	1083170.7
1005	5580	130093.2

	metrics.statistics.words
0	121533
1	6067
2	26389
3	74959
4	22022
...	...
1001	6913
1002	9153
1003	191182
1004	730715
1005	95496

[1006 rows x 38 columns]

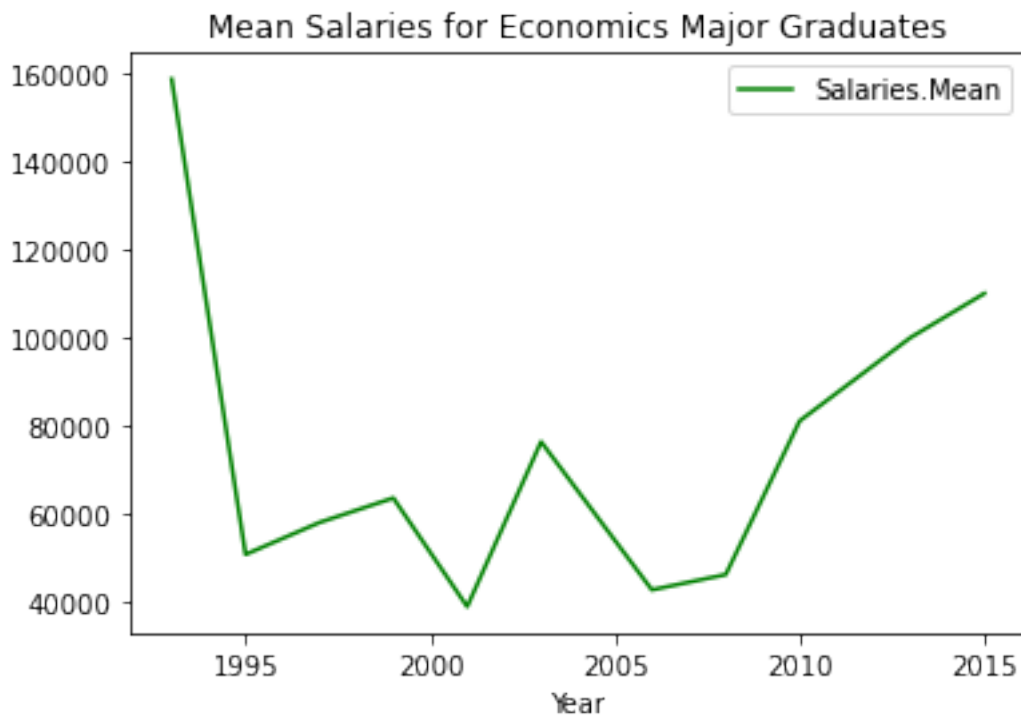
Using dataframes returned by your new function, make a line plot, a bar plot, and a histogram plot

- Each plot should be of a different CORGIS dataset

- 3 plots total (not 3 plots for each dataset)
- Note that when you make these plots, there should not be any data files in your notebook's directory.
- You may find it useful to use `df.head()` to get a quick idea of what's in each dataframe.
- Remember that bar plots have categorical data on the horizontal axis, and depending on your data, you may want to subset your data before making the bar plot.

```
[10]: #line plot
graduates = pdcorgis("Graduates")
graduates.head()
graduates_economics = graduates[graduates["Education.Major"] == "Economics"]
graduates_economics.plot(kind='line', x="Year", y="Salaries.Mean", color = "green", title = 'Mean Salaries for Economics Major Graduates')
```

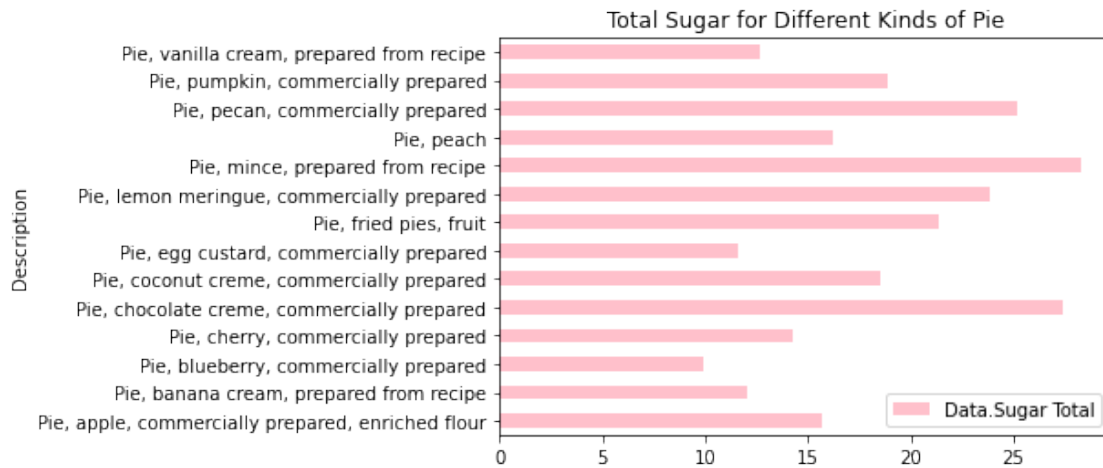
```
[10]: <AxesSubplot:title={'center':'Mean Salaries for Economics Major Graduates'},
xlabel='Year'>
```



```
[11]: #bar plot
ingredients = pdcorgis("Ingredients")
ingredients.head()
ingredients_pie = ingredients[ingredients["Category"] == "Pie"]
ingredients_pie.plot(kind='barh', x="Description", y="Data.Sugar Total", color = "pink", title = 'Total Sugar for Different Kinds of Pie')
```



```
[11]: <AxesSubplot:title={'center':'Total Sugar for Different Kinds of Pie'},
      ylabel='Description'>
```



```
[12]: #histogram plot
broadway = pdcorgis("Broadway")
broadway.head()
beauty_and_the_beast = Broadway[Broadway["Show.Name"] == "Beauty And The Beast"]
beauty_and_the_beast["Statistics.Capacity"].plot.hist(bins = 10, color = "#483D8B",
    →"darkviolet", title = 'Distribution of Theatre Capacity Across Beauty and
    →the Beast Performances')
```

```
[12]: <AxesSubplot:title={'center':'Distribution of Theatre Capacity Across Beauty and
      the Beast Performances'}, ylabel='Frequency'>
```

Distribution of Theatre Capacity Across Beauty and the Beast Performances

