elastic{on} TOUR

Elasticsearchを活用して、統計データから新たなインサイトを生成する取り組み

Jesus Jackson, Chief Data Scientist

Daewoo Chong, Lead Data Scientist

ブーズ·アレン·ハミルトン

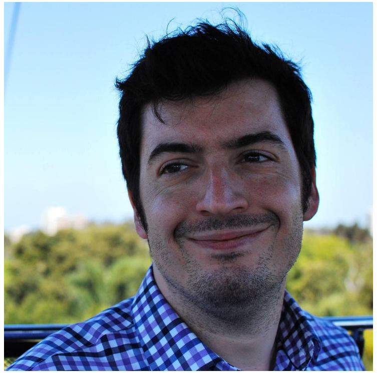
2016年10月13日

United States™ Census Bureau

アジェンダ

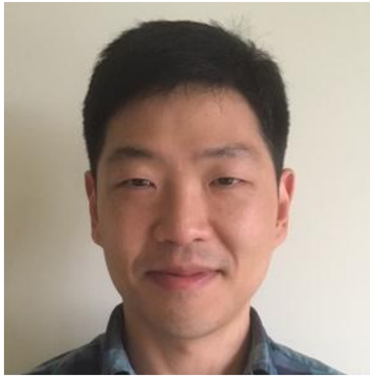
- はじめに
- 検索プロトタイプ
  - プロトタイプとは？
  - 統計データについて
- インデキシングストラテジー
  - 統計データのインデキシングにおける課題とは？
  - 候補となるストラテジー現行のストラテジー
- プロトタイプの機能
  - プロトタイプの主要機能
  - Elasticsearchの主要機能
- デモ
- ご質問をどうぞ！

はじめに



Jesus Jackson

- ブーズ･アレン･ハミルトンに7年+勤務
- 金融サービス業界向け連邦データサイエンスサービス部門責任者
- 自身について一言：先日スカイダイビングしました



Daewoo Chong

- ブーズ･アレン･ハミルトン1.5年+勤務
- 機械学習マニア
- 自身について一言：オフの日はニューラルネットを作成してます

他のチームメンバー：

Julia Stevens
Raj Cheekatamarla
Ram Anusuri

クライアントチームメンバー：

Richie Wang
David MacCormack
Zachary Whitman

検索プロトタイプ

プロトタイプとは？

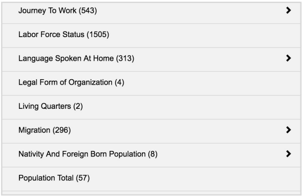
ユーザーがシンプルに、いつでも妥当なデータにアクセスできる機能を実現する

elastic

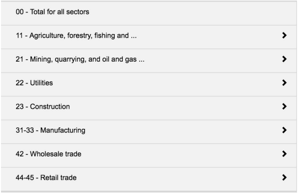
検索プロトタイプ

統計データについて

- 統計データは非常に情報量が多く、複雑。たとえば：
  - 調査/プログラムの種類：米国コミュニティ調査、国勢調査、事業主調査など
  - 地理的区分：地域、州、郡、学区、細分区など
  - トピック：年齢、コンピューターとインターネットの使用状況、貧困度、人種、性別など
  - 業界：建設、金融・保険、製造、卸売など

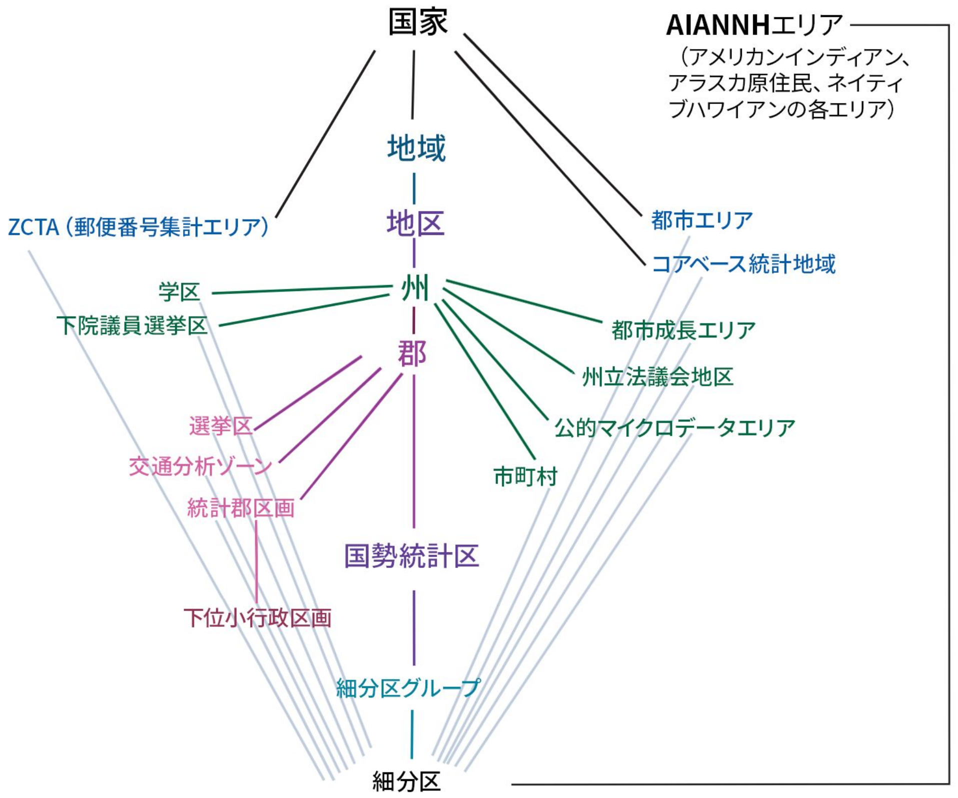


Journey To Work (543) ›
Labor Force Status (1505)
Language Spoken At Home (313) ›
Legal Form of Organization (4)
Living Quarters (2)
Migration (296) ›
Nativity And Foreign Born Population (8) ›
Population Total (57)



00 - Total for all sectors
11 - Agriculture, forestry, fishing and ... ›
21 - Mining, quarrying, and oil and gas ... ›
22 - Utilities ›
23 - Construction ›
31-33 - Manufacturing ›
42 - Wholesale trade ›
44-45 - Retail trade ›

検索プロトタイプ : *地理階層*



国家

地域

地区

州

郡

国勢統計区

細分区グループ

細分区

AIANNHエリア
（アメリカンインディアン、
アラスカ原住民、ネイティ
ブハワイアンの各エリア）

ZCTA（郵便番号集計エリア）

都市エリア

コアベース統計地域

学区

下院議員選挙区

都市成長エリア

州立法議会地区

公的マイクロデータエリア

市町村

選挙区

交通分析ゾーン

統計郡区画

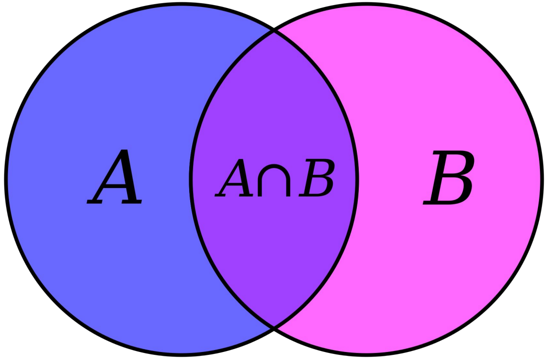
下位小行政区画

インデキシングストラテジー

インデキシングストラテジー

統計データのインデキシングにおける課題とは？

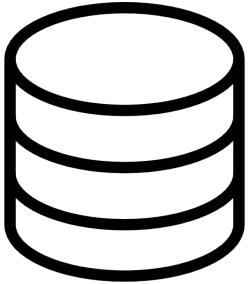
- ブール検索モデル
- トピック、地理、業界、データセット、調査年を横断するフィルタリング
- デッドエンドなし
- 応答時間100ミリ秒以下



インデキシングストラテジー

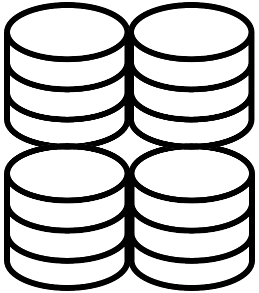
候補となるストラテジー現行のストラテジー

非正規化



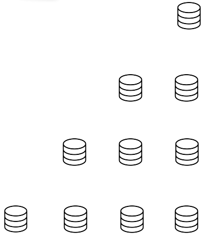
クエリしやすいが
サイズが大きくコスト高

ペアレント/チャイルド



小サイズでコストは
高くないが応答時間
が遅い

準正規化



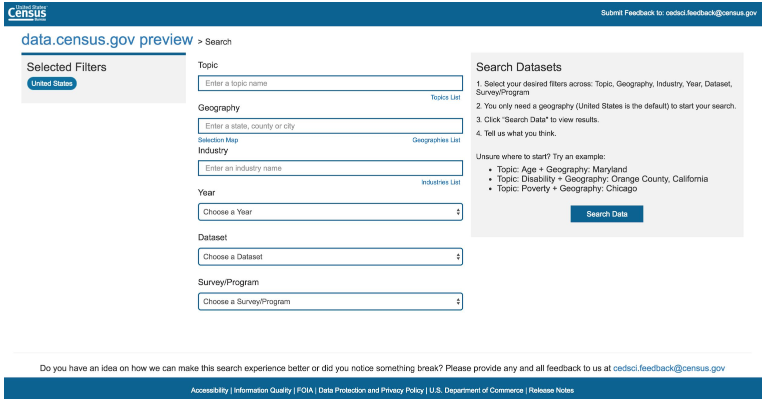
小サイズで経済的、応
答時間も早いが、セル
値を持たない

プロトタイプの機能

プロトタイプの機能

プロトタイプの主要機能

- デッドエンドなし
- 応答時間100ミリ秒以下
- ログのパース/インデックス/分析にElastic Stackを使用
- 認証とアクセスベースのロール制御に Shieldを使用


United States Census Bureau
Submit Feedback to: cedsci.feedback@census.gov
Topics List
Selection Map
Geographies List
Industries List
Search Data
Accessibility | Information Quality | FOIA | Data Protection and Privacy Policy | U.S. Department of Commerce | Release Notes

data.census.gov preview > Search

Selected Filters

United States

Topic

Enter a topic name

Geography

Enter a state, county or city

Industry

Enter an industry name

Year

Choose a Year

Dataset

Choose a Dataset

Survey/Program

Choose a Survey/Program

Search Datasets

1. Select your desired filters across: Topic, Geography, Industry, Year, Dataset, Survey/Program
2. You only need a geography (United States is the default) to start your search.
3. Click "Search Data" to view results.
4. Tell us what you think.

Unsure where to start? Try an example:

- Topic: Age + Geography: Maryland
- Topic: Disability + Geography: Orange County, California
- Topic: Poverty + Geography: Chicago

Do you have an idea on how we can make this search experience better or did you notice something break? Please provide any and all feedback to us at cedsci.feedback@census.gov

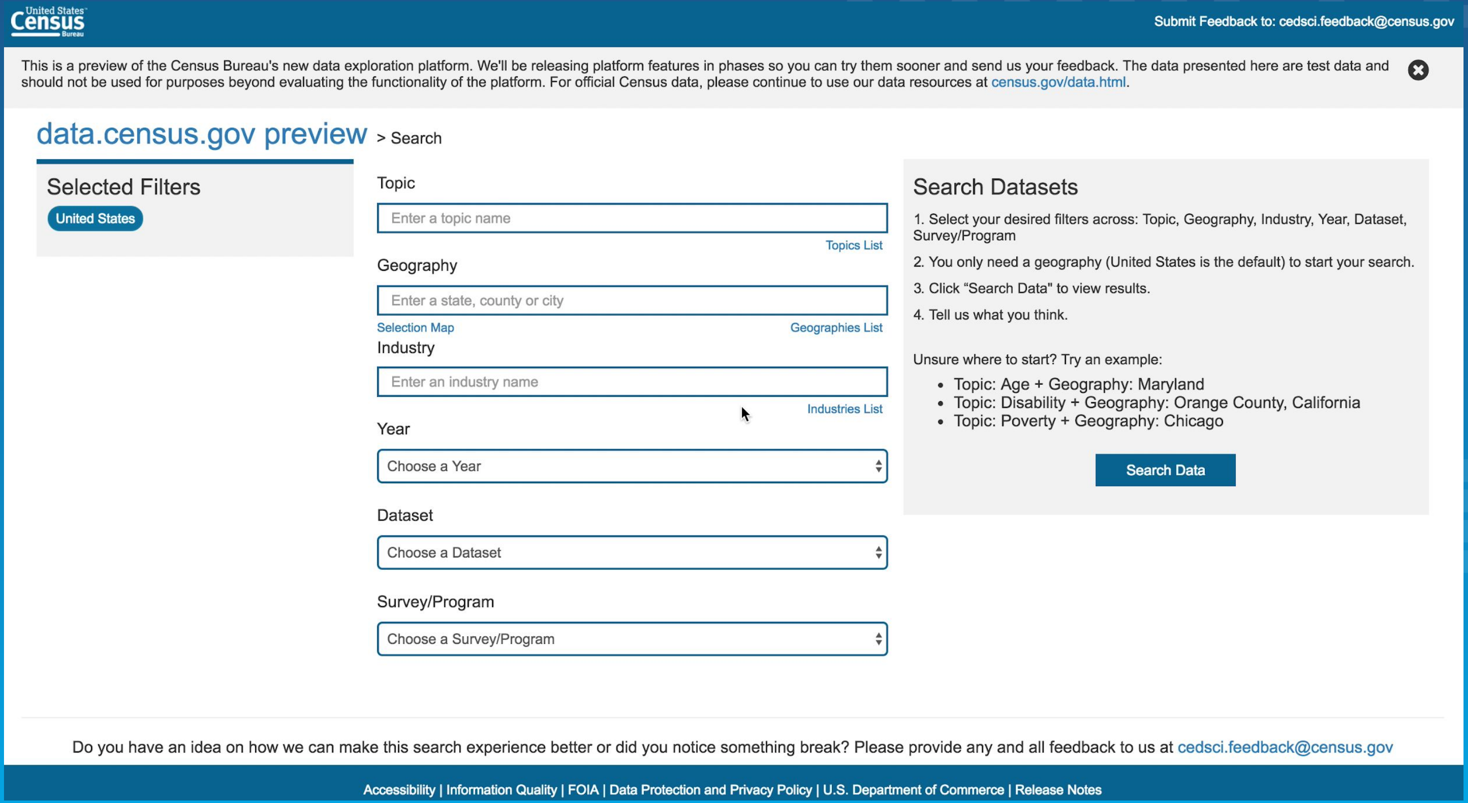
プロトタイプの機能

プロトタイプにメリットをもたらした主要な Elasticsearchの機能

- クエリキャッシングで応答時間が最大 25%向上
- マッピングAPI
- 関連性モデル（例：TF/IDF、BM25）
- クエリAPI
- 複数言語のサポート（例：JavaScript、Python、Java）

デモ

elastic



Submit Feedback to: cedsci.feedback@census.gov

This is a preview of the Census Bureau's new data exploration platform. We'll be releasing platform features in phases so you can try them sooner and send us your feedback. The data presented here are test data and should not be used for purposes beyond evaluating the functionality of the platform. For official Census data, please continue to use our data resources at census.gov/data.html.

data.census.gov preview > Search

Selected Filters

United States

Topic

Enter a topic name

Topics List

Geography

Enter a state, county or city

Selection Map | Geographies List

Industry

Enter an industry name

Industries List

Year

Choose a Year

Dataset

Choose a Dataset

Survey/Program

Choose a Survey/Program

Search Datasets

1. Select your desired filters across: Topic, Geography, Industry, Year, Dataset, Survey/Program
2. You only need a geography (United States is the default) to start your search.
3. Click "Search Data" to view results.
4. Tell us what you think.

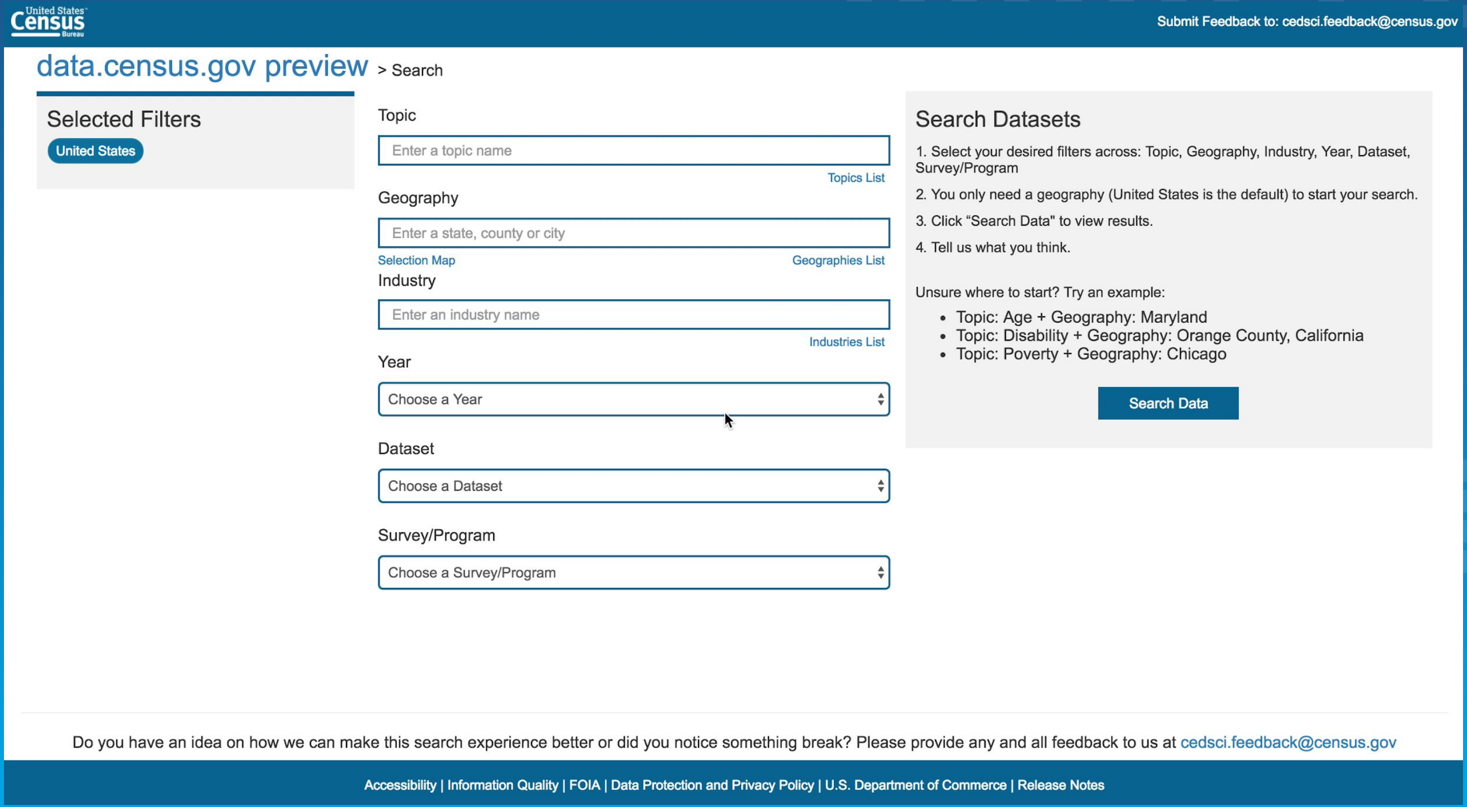
Unsure where to start? Try an example:

- Topic: Age + Geography: Maryland
- Topic: Disability + Geography: Orange County, California
- Topic: Poverty + Geography: Chicago

Search Data

Do you have an idea on how we can make this search experience better or did you notice something break? Please provide any and all feedback to us at cedsci.feedback@census.gov

Accessibility | Information Quality | FOIA | Data Protection and Privacy Policy | U.S. Department of Commerce | Release Notes


United States Census Bureau
Submit Feedback to: cedsci.feedback@census.gov
data.census.gov preview > Search
Selected Filters
United States
Topic
Enter a topic name
Topics List
Geography
Enter a state, county or city
Selection Map
Geographies List
Industry
Enter an industry name
Industries List
Year
Choose a Year
Dataset
Choose a Dataset
Survey/Program
Choose a Survey/Program
Search Data
Accessibility | Information Quality | FOIA | Data Protection and Privacy Policy | U.S. Department of Commerce | Release Notes

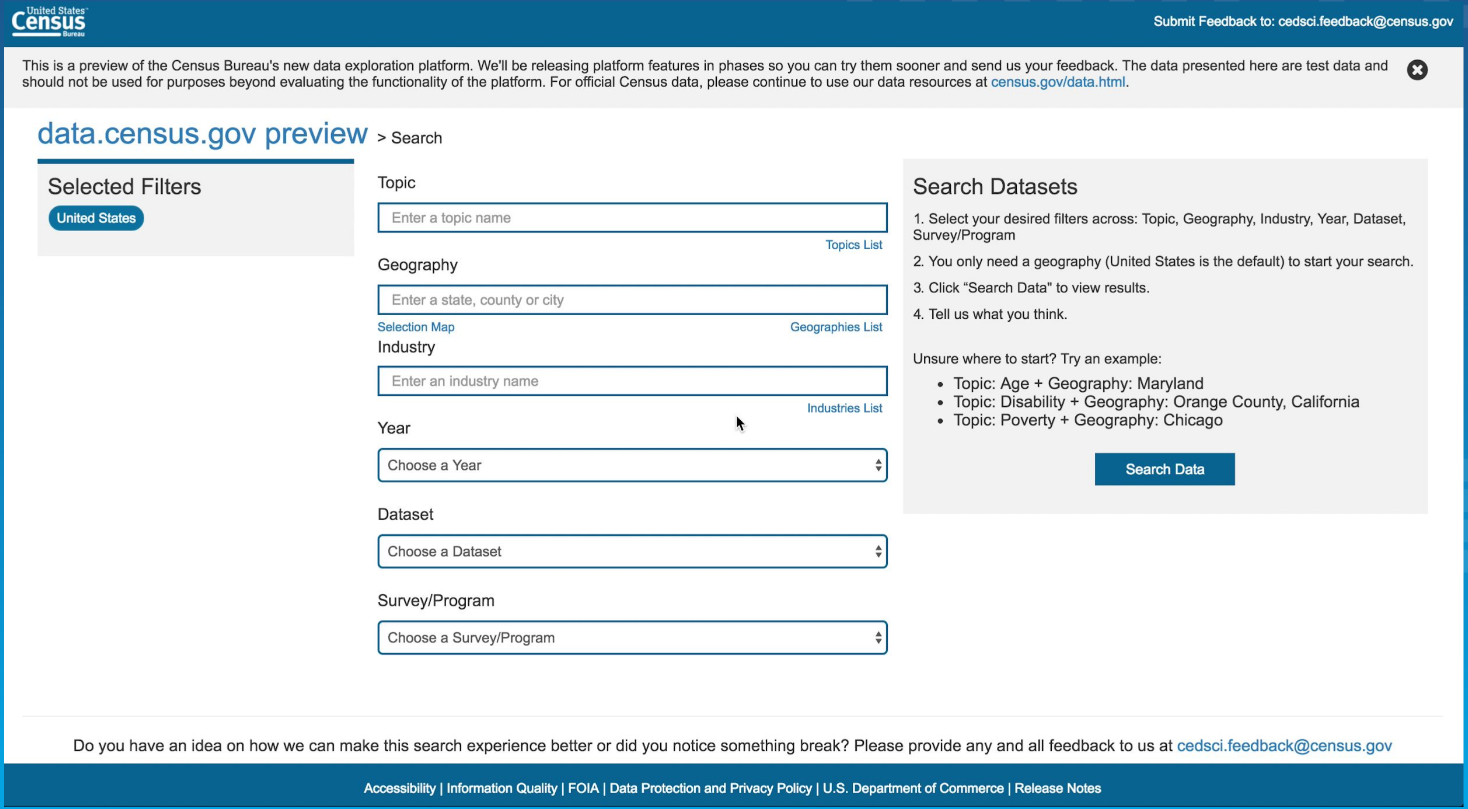
Search Datasets

1. Select your desired filters across: Topic, Geography, Industry, Year, Dataset, Survey/Program
2. You only need a geography (United States is the default) to start your search.
3. Click "Search Data" to view results.
4. Tell us what you think.

Unsure where to start? Try an example:

- Topic: Age + Geography: Maryland
- Topic: Disability + Geography: Orange County, California
- Topic: Poverty + Geography: Chicago

Do you have an idea on how we can make this search experience better or did you notice something break? Please provide any and all feedback to us at cedsci.feedback@census.gov



Submit Feedback to: cedsci.feedback@census.gov

This is a preview of the Census Bureau's new data exploration platform. We'll be releasing platform features in phases so you can try them sooner and send us your feedback. The data presented here are test data and should not be used for purposes beyond evaluating the functionality of the platform. For official Census data, please continue to use our data resources at census.gov/data.html.

data.census.gov preview > Search

Selected Filters

United States

Topic: Enter a topic name

Topics List

Geography: Enter a state, county or city

Selection Map

Geographies List

Industry: Enter an industry name

Industries List

Year: Choose a Year

Dataset: Choose a Dataset

Survey/Program: Choose a Survey/Program

Search Datasets

1. Select your desired filters across: Topic, Geography, Industry, Year, Dataset, Survey/Program
2. You only need a geography (United States is the default) to start your search.
3. Click “Search Data" to view results.
4. Tell us what you think.

Unsure where to start? Try an example:

- Topic: Age + Geography: Maryland
- Topic: Disability + Geography: Orange County, California
- Topic: Poverty + Geography: Chicago

Search Data

Do you have an idea on how we can make this search experience better or did you notice something break? Please provide any and all feedback to us at cedsci.feedback@census.gov

Accessibility | Information Quality | FOIA | Data Protection and Privacy Policy | U.S. Department of Commerce | Release Notes

elastic

ご質問をどうぞ！

elastic

elastic

www.elastic.co

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nd/4.0/

Creative Commons and the double C in a circle are
registered trademarks of Creative Commons in the United States and other countries.
Third party marks and brands are the property of their respective holders.