

国勢調査年齢別人口の きれいなデータ

林 正洋

作ろうと思ったきっかけ

国勢調査の人口の時系列比較を
するアプリを作りたい

e-Statの国勢調査APIデータ

2010年調査

時間軸（年次）	表章項目	地域（2010）	国籍2010	男女別2010	年齢4	単位	値
2010000000：2010年	020：人口	01100：札幌市	000：総数（国籍）	000：総数（男女別）	000：総数（年齢）	人	1913545
2010000000：2010年	020：人口	01101：中央区	000：総数（国籍）	000：総数（男女別）	000：総数（年齢）	人	220189
2010000000：2010年	020：人口	01102：北区	000：総数（国籍）	000：総数（男女別）	000：総数（年齢）	人	278781
2010000000：2010年	020：人口	01103：東区	000：総数（国籍）	000：総数（男女別）	000：総数（年齢）	人	255873

2015年調査

時間軸（年次）	表章項目	地域（2015）	国籍_2015	男女別_2015	年齢_2015	単位	値
2015000000：2015年	020：人口	01100：札幌市	0000：総数（国籍）	0000：総数（男女別）	0000：総数（年齢）	人	1952356
2015000000：2015年	020：人口	01101：札幌市 中央区	0000：総数（国籍）	0000：総数（男女別）	0000：総数（年齢）	人	237627
2015000000：2015年	020：人口	01102：札幌市 北区	0000：総数（国籍）	0000：総数（男女別）	0000：総数（年齢）	人	285321
2015000000：2015年	020：人口	01103：札幌市 東区	0000：総数（国籍）	0000：総数（男女別）	0000：総数（年齢）	人	261912

2020年調査

時間軸（年次）	表章事項	全国，都道府県，市区町村	国籍総数が日本人	男女	年齢	単位	値
2020000000：2020年	2020_01：人口	01100：札幌市	0：国籍総数	0：総数	000：総数	人	1973395
2020000000：2020年	2020_01：人口	01101：札幌市中央区	0：国籍総数	0：総数	000：総数	人	248680
2020000000：2020年	2020_01：人口	01102：札幌市北区	0：国籍総数	0：総数	000：総数	人	289323
2020000000：2020年	2020_01：人口	01103：札幌市東区	0：国籍総数	0：総数	000：総数	人	265379

同じように見えるが年ごとにいろいろ違う...

例1) 男女の別を示す項目

2010年調査

男女別2010	
000	: 総数 (男女別)
000	: 総数 (男女別)
000	: 総数 (男女別)
000	: 総数 (男女別)

2015年調査

男女別_2015	
0000	: 総数 (男女別)
0000	: 総数 (男女別)
0000	: 総数 (男女別)
0000	: 総数 (男女別)

2020年調査

男女	
0	: 総数
0	: 総数
0	: 総数
0	: 総数

例2) 集計地域を示す項目

2010年調査

地域 (2010)
01100 : 札幌市
01101 : 中央区
01102 : 北区
01103 : 東区

2015年調査

地域 (2015)
01100 : 札幌市
01101 : 札幌市 中央区
01102 : 札幌市 北区
01103 : 札幌市 東区

半角スペースあり

2020年調査

全国, 都道府県, 市区町村
01100 : 札幌市
01101 : 札幌市中央区
01102 : 札幌市北区
01103 : 札幌市東区

半角スペースなし

作品：国勢調査 年齢別人口のきれいなデータ

- * 中身はe-Statから取ってきた国勢調査データ(1980年～2020年)
- * データの形式をすべて揃える加工を行ったもの

管理番号	年次	表章項目	地域コード	地域	上位都道府県コード	上位都道府県	全域・人口集中地区	国籍	男女別	年齢	単位	人口
Management Number	Survey year	Represents	Area code	Area	Pref code	Prefectures	Whole or DID	Nationality	Sex	Age	Unit	Population
mngmtNo	surveyYear	represents	areaCode	area	prefCode	pref	wholeOrDID	nationality	sex	Age	unit	population
202000000928	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	000：総数	0000：総数	人	248,680
202000000929	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	001：男	0000：総数	人	112,853
202000000930	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	002：女	0000：総数	人	135,827
202000000931	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	000：総数	1000：0歳	人	1,490
202000000932	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	001：男	1000：0歳	人	718
202000000933	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	002：女	1000：0歳	人	772
202000000934	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	000：総数	1001：1歳	人	1,559
202000000935	2020	000：人口	01101	札幌市中央区	01000	北海道	000：全域	000：総数	001：男	1001：1歳	人	832

※これはCSV版

このデータの特徴

- * 「分析のためのデータ」としての価値が高い
特に時系列推移をみる分析に有用
- * リレーショナルデータベースにそのままつっこめる
SQLで様々な抽出が可能

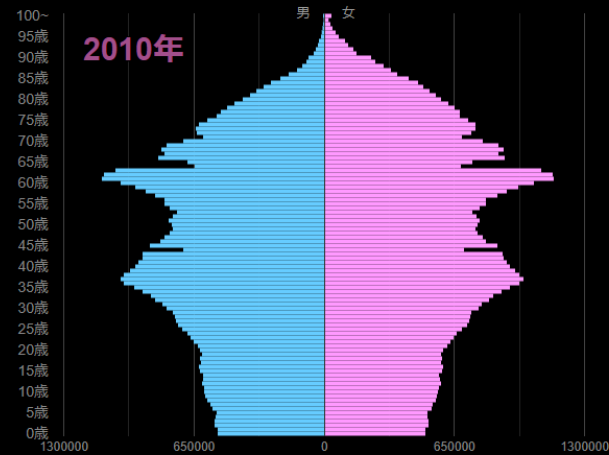
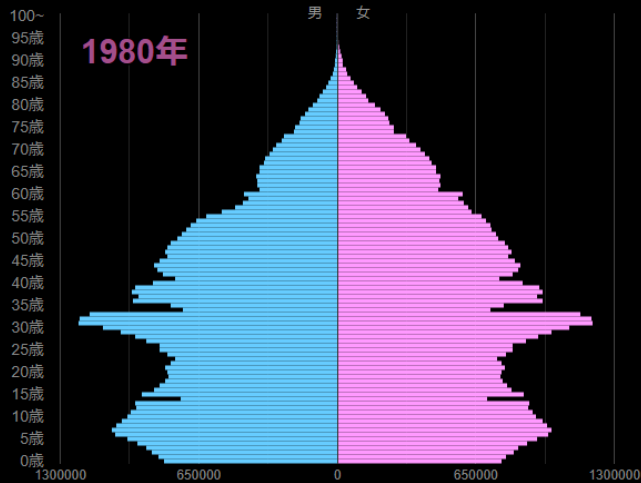
既存のデータをクレンジングし
利便性を向上させることでの
データの価値向上

このようなアプリの作成にとっても便利です

全国人口ピラミッド

2020年国勢調査対応版

日本 (全国)



1980 1985 1990 1995 2000 2005 2010 2015 2020 ループ 色変更

1980 1985 1990 1995 2000 2005 2010 2015 2020 ループ 色変更

<https://www.mirko.jp/pyramid/>

今回作ったRDF版データ

```
:K202000000928 a qb:Observation ;
  cd-dimension:timePeriod "2020"^^xsd:gYear ;
  dcterms:identifier "01101" ;
  sdmx-dimension:refArea "札幌市中央区"@ja ;
  dcterms:isPartOf "01000" ;
  sacs:prefectureLabel "北海道"@ja ;
  g00200521-dimension-2010:area g00200521-code-2010:area-all ;
  cd-dimension:nationality cd-code:nationality-all ;
  cd-dimension:sex cd-code:sex-all ;
  cd-dimension:age cd-code:age-all ;
  estat-measure:population "248680"^^xsd:decimal .

:K202000000929 a qb:Observation ;
  cd-dimension:timePeriod "2020"^^xsd:gYear ;
  dcterms:identifier "01101" ;
  sdmx-dimension:refArea "札幌市中央区"@ja ;
  dcterms:isPartOf "01000" ;
  sacs:prefectureLabel "北海道"@ja ;
  g00200521-dimension-2010:area g00200521-code-2010:area-all ;
  cd-dimension:nationality cd-code:nationality-all ;
  cd-dimension:sex cd-code:sex-male ;
  cd-dimension:age cd-code:age-all ;
  estat-measure:population "112853"^^xsd:decimal .
```

Oracle Cloud の RDF Graph Server

(無料版)



ORACLE

Special thanks!
日本オラクル株式会社
の皆さま

SPARQLテストフォーム公開中

https://www.mirko.jp/pyramid/ora_db/sp.html

サンプルクエリもいろいろ作りました

全国＋各都道府県／国籍総数／男性／年齢総数の人口を自治体コード順に並べる

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>
PREFIX cd-dimension: <http://data.e-stat.go.jp/lod/ontology/crossDomain/dimension/>
PREFIX g00200521-code-2010: <http://data.e-stat.go.jp/lod/ontology/g00200521/code/2010/>
PREFIX cd-code: <http://data.e-stat.go.jp/lod/ontology/crossDomain/code/>
PREFIX sacs: <http://data.e-stat.go.jp/lod/terms/sacs#>
PREFIX estat-measure: <http://data.e-stat.go.jp/lod/ontology/measure/>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?areacode ?areaname ?population
WHERE {
  ?s cd-dimension:timePeriod "2020"^^xsd:gYear ;
    cd-dimension:nationality cd-code:nationality-all ;
    cd-dimension:sex cd-code:sex-male ;
    cd-dimension:age cd-code:age-all ;
    dcterms:identifier ?areacode ;
    dcterms:isPartOf "-";
    sdmx-dimension:refArea ?areaname ;
    estat-measure:population ?population .
}
ORDER BY ?areacode
```

京都市左京区／国籍総数／男女総数 の年齢別人口

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>
PREFIX cd-dimension: <http://data.e-stat.go.jp/lod/ontology/crossDomain/dimension/>
PREFIX g00200521-code-2010: <http://data.e-stat.go.jp/lod/ontology/g00200521/code/2010/>
PREFIX cd-code: <http://data.e-stat.go.jp/lod/ontology/crossDomain/code/>
PREFIX sacs: <http://data.e-stat.go.jp/lod/terms/sacs#>
PREFIX estat-measure: <http://data.e-stat.go.jp/lod/ontology/measure/>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?age ?population
WHERE {
  ?s cd-dimension:timePeriod "2020"^^xsd:gYear ;
    sdmx-dimension:refArea "京都市左京区"@ja ;
    cd-dimension:nationality cd-code:nationality-all ;
    cd-dimension:sex cd-code:sex-all ;
    cd-dimension:age ?age ;
    dcterms:identifier ?areacode ;
    estat-measure:population ?population .
}
ORDER BY ?s
```

2020年国勢調査／神奈川県／日本人のみ／女性／35歳の人口

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>
PREFIX cd-dimension: <http://data.e-stat.go.jp/lod/ontology/crossDomain/dimension/>
PREFIX g00200521-code-2010: <http://data.e-stat.go.jp/lod/ontology/g00200521/code/2010/>
PREFIX cd-code: <http://data.e-stat.go.jp/lod/ontology/crossDomain/code/>
PREFIX sacs: <http://data.e-stat.go.jp/lod/terms/sacs#>
PREFIX estat-measure: <http://data.e-stat.go.jp/lod/ontology/measure/>
SELECT ?population
WHERE {
  ?s cd-dimension:timePeriod "2020"^^xsd:gYear ;
    sdmx-dimension:refArea "神奈川県"@ja ;
    cd-dimension:nationality cd-code:nationality-japan ;
    cd-dimension:sex cd-code:sex-female ;
    cd-dimension:age cd-code:age-35 ;
    estat-measure:population ?population .
}
```

市区町村の人口ランキング

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#>
PREFIX cd-dimension: <http://data.e-stat.go.jp/lod/ontology/crossDomain/dimension/>
PREFIX g00200521-code-2010: <http://data.e-stat.go.jp/lod/ontology/g00200521/code/2010/>
PREFIX cd-code: <http://data.e-stat.go.jp/lod/ontology/crossDomain/code/>
PREFIX sacs: <http://data.e-stat.go.jp/lod/terms/sacs#>
PREFIX estat-measure: <http://data.e-stat.go.jp/lod/ontology/measure/>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?areaname ?population
WHERE {
  ?s cd-dimension:timePeriod "2020"^^xsd:gYear ;
    cd-dimension:nationality cd-code:nationality-all ;
    cd-dimension:sex cd-code:sex-all ;
    cd-dimension:age cd-code:age-all ;
    sdmx-dimension:refArea ?areaname ;
    estat-measure:population ?population .
  MINUS {?s dcterms:isPartOf "-".}
}
ORDER BY DESC (?population)
LIMIT 100
```

国勢調査年齢別人口の きれいなデータ

<https://www.mirko.jp/pyramid/data/>