

Search and Download datasets from data.gov.in

Sainath Adapa

25 March 2017

What is data.gov.in

- Open Government Data (OGD) Platform India - data.gov.in - is a platform for supporting Open Data initiative of Government of India.
- The portal is intended to be used by Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use.
- It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective.

Searching and Downloading a dataset from data.gov.in

1. Enter keywords and search
2. Click on the relevant search result. This will take you to the catalog containing that dataset.
3. Go through pages in the catalog, to find the right dataset
4. Download the dataset

Step 1 - Enter keywords and search



data.gov 
Open Government Data (OGD) Platform India

age sex population



LOG IN | REG

Step 2 - Click on the relevant search result

age sex population

SEARCH

Search results

Population Classified By Place Of Birth, Age And Sex, Census 2001 - India And States

The catalog contains data related to population classified by place of birth, age and sex, census 2001 - India and States. It includes data on Population by Age, Population by Age Group, Population Classified by Age, Population by Place of Birth, Population by Sex, Population Classified by Sex, Population by Place of Birth and Sex, Population Classified by Place of Birth and Age, Census 2001.

Population In Five Year Age-Group By Residence And Sex, Census 2001 - India And States

The catalog contains data related to population in five year age-group by residence and sex, census 2001 - India and states. It includes data on Population in Five Year Age-group, Population in Five Year Age-group by Residence, Population in Five Year Age-group by Sex , Five Year Age-group Population, Census 2001, Scheduled Caste (SC) and Scheduled Tribe (ST).

Projected Population Characteristics

The data refers to Projected population characteristics(Population,Sex ratio,Population density,Population by broad age-group,Proportion (percent),Median age,Dependency Ratio,Population growth rate,Crude Birth rate(CBR),Crude Death rate (CDR),Infant Mortality Rate (IMR),Under-5 mortality rate,Total Fertility Rate (TFR),Life expectancy of males,Life expectancy of females) of India.

Single Year Age Returns By Residence And Sex, Census 2001 - India And States

The catalog contains data related to single year age returns by residence and sex, census 2001 - India and states. It includes data on Single Year Age Returns, Single Year Age Returns by Residence, Single Year Age Returns by Sex, Single Year Age Returns Population, Census 2001, Scheduled Caste (SC) and Scheduled Tribe (ST).

Step 3 - Find the dataset in the catalog

🏠 / Catalogs / Population Classified by Place of Birth, Age and Sex, Census 2001 - India and States



Catalog Info

The catalog contains data related to population classified by place of birth, age and sex, census 2001 - India and States. It includes data on Population by Age, Population by Age Group, Population Classified by Age, Population by Place of Birth, Population b... [< Show more >](#)

- Released Under: [National Data Sharing and Accessibility Policy \(NDSAP\)](#)
- Contributor: [Ministry of Home Affairs](#)
Department of Home
Registrar General and Census Commissioner,
India
- Keywords: [Population](#) [Age](#) [age group](#)
[Place of Birth](#) [Sex](#) [Census 2001](#)
[census](#)
- Group: [Census Digital Library](#)
- Sectors: [Population Enumerator](#)
- Published on Data Portal: August 17, 2016



Chief Data Officer

36 Resource(s) found

Title

Format

APPLY



Population Classified by Place of Birth, Age and Sex, 2001 - India



Visual Access: NA



Granularity: Decadal



File Size: 2.37 MB



Download: 216



Reference URL:
<http://www.censusindia.gov.in/DigitalLib...>



Request API: 1



Note: NA



Population Classified by Place of Birth, Age and Sex, 2001 - Sikkim



Visual Access: NA



Granularity: Decadal



File Size: 358.5 KB



Download: 78



Reference URL:
<http://www.censusindia.gov.in/DigitalLib...>



Request API: 0



Note: NA

Step 3 - Find the dataset in the catalog

If you are lucky, you will find the dataset that you are looking for, on the first page

 Population Classified by Place of Birth, Age and Sex, 2001 - Bihar

 EXCEL

 Visual Access: NA

 Granularity: Decadal

 File Size: 2.49 MB

 Download: 79

 Reference URL:
<http://www.censusindia.gov.in/DigitalLib...>

 Request API:  0

 Note: NA

 Population Classified by Place of Birth, Age and Sex, 2001 - Delhi

 EXCEL

 Visual Access: NA

 Granularity: Decadal

 File Size: 690.5 KB

 Download: 85

 Reference URL:
<http://www.censusindia.gov.in/DigitalLib...>

 Request API:  0

 Note: NA

123456next>last»

ogdindiar package

- OGD (Open Government Data) INDIA R
- Available at <https://github.com/steadyfish/ogdindiar>
- Not on CRAN yet. Install using the command `devtools::install_github("steadyfish/ogdindiar")`
- Provides functions to search and download datasets from data.gov.in
- Since there is no API available for searching datasets, these functions use `rvest` to do web scraping
- Also, ogdindiar provides a function (`fetch_data`) to download a dataset using the API. Note that not all datasets have API access.
- Refer to the vignettes for more information
 1. [API access](#)
 2. [Search functionality](#)

Search for the right catalog

```
library(dplyr)
library(ogdindiar)

catalogs_df <- search_for_datasets(search_terms      = 'age sex population',
                                   limit_catalog_pages = Inf,
                                   limit_catalogs      = 10,
                                   return_catalog_list = TRUE)
```

```
catalogs_df %>% glimpse
```

```
## Observations: 18
## Variables: 2
## $ name <chr> "Population Classified by Place of Birth, Age and Sex, Ce...
## $ link <chr> "https://data.gov.in/catalog/population-classified-place-...
```

Search for the right catalog

```
catalogs_df$name
```

```
## [1] "Population Classified by Place of Birth, Age and Sex, Census 2001 - India and States"
## [2] "Population in Five Year Age-group by Residence and Sex, Census 2001 - India and States"
## [3] "Projected population characteristics"
## [4] "Single Year Age Returns by Residence and Sex, Census 2001 - India and States"
## [5] "Educational Level by Age and Sex for Population Age 7 and Above, 2011 - India and States"
## [6] "Population attending educational institution by age, sex and type of educational institution - India and States"
## [7] "Prison inmate population by sex and age-group"
## [8] "Disabled Population by Type of Disability, Age and Sex, Census 2001 - India and States"
## [9] "Population by Bilingualism, Trilingualism, Age and Sex, Census 2001 - India and States"
## [10] "Main Workers Classified by Industrial Category, Age and Sex, Census 1991 - India and States"
## [11] "Disabled Population by Type of Disability, Marital Status, Age and Sex, Census 2001 - India and States"
## [12] "Disabled Population among Main Workers, Marginal Workers, Non-Workers by Type of Disability, Age and Sex, Census 2001"
## [13] "Education Level Graduate and Above by Sex for Population Age 15 and Above - India and States"
## [14] "Marital Status by Single Year Age and Sex - India and States"
## [15] "Population Age 5-19 Attending School/ College by Economic Activity Status and Sex (For Each Caste/Tribe Separately), "
## [16] "Age, Sex and Educational level Population, Census 1991 - India and States"
## [17] "Population ages 5-19 attending educational institutions by economic activity status and sex - India and States"
## [18] "Education Level by Age and Sex for Population Age 7 and Above, Census 2001 - India and States"
```

Get the list of datasets from the catalog

```
datasets_df <- get_datasets_from_a_catalog(catalog_link = catalogs_df$link[1],
                                           limit_dataset_pages = Inf,
                                           limit_datasets = 5)
```

```
datasets_df %>% glimpse
```

```
## Observations: 6
## Variables: 13
## $ name      <chr> "Population Classified by Place of Birth, Age and ...
## $ granularity <chr> "Decadal", "Decadal", "Decadal", "Decadal", "Decad...
## $ file_size  <chr> "2.37 MB", "358.5 KB", "4.63 MB", "1022.5 KB", "2....
## $ downloads  <dbl> 217, 78, 96, 86, 79, 87
## $ res_id     <chr> NA, NA, NA, NA, NA, NA
## $ default    <chr> "https://data.gov.in/resources/population-classifi...
## $ csv        <chr> NA, NA, NA, NA, NA, NA
## $ excel      <chr> "https://data.gov.in/resources/population-classifi...
## $ ods        <chr> NA, NA, NA, NA, NA, NA
## $ xls        <chr> NA, NA, NA, NA, NA, NA
## $ json       <chr> NA, NA, NA, NA, NA, NA
## $ xml        <chr> NA, NA, NA, NA, NA, NA
## $ jsonp      <chr> NA, NA, NA, NA, NA, NA
```

Get the list of datasets from the catalog

```
datasets_df$name
```

```
## [1] "Population Classified by Place of Birth, Age and Sex, 2001 - India"  
## [2] "Population Classified by Place of Birth, Age and Sex, 2001 - Sikkim"  
## [3] "Population Classified by Place of Birth, Age and Sex, 2001 - Uttar Pradesh"  
## [4] "Population Classified by Place of Birth, Age and Sex, 2001 - Jammu & Kashmir"  
## [5] "Population Classified by Place of Birth, Age and Sex, 2001 - Bihar"  
## [6] "Population Classified by Place of Birth, Age and Sex, 2001 - Delhi"
```

Download a dataset

```
datasets_df %>% slice(6) %>% glimpse
```

```
## Observations: 1
## Variables: 13
## $ name      <chr> "Population Classified by Place of Birth, Age and ..."
## $ granularity <chr> "Decadal"
## $ file_size  <chr> "690.5 KB"
## $ downloads  <dbl> 87
## $ res_id     <chr> NA
## $ default    <chr> "https://data.gov.in/resources/population-classifi..."
## $ csv        <chr> NA
## $ excel      <chr> "https://data.gov.in/resources/population-classifi..."
## $ ods        <chr> NA
## $ xls        <chr> NA
## $ json       <chr> NA
## $ xml        <chr> NA
## $ jsonp      <chr> NA
```

```
download_dataset(urllink = datasets_df$excel[6], filepath = 'delhi.xls')
```

```
## $filepath
## [1] "delhi.xls"
##
## $type
## [1] "xls"
```

Reading the dataset

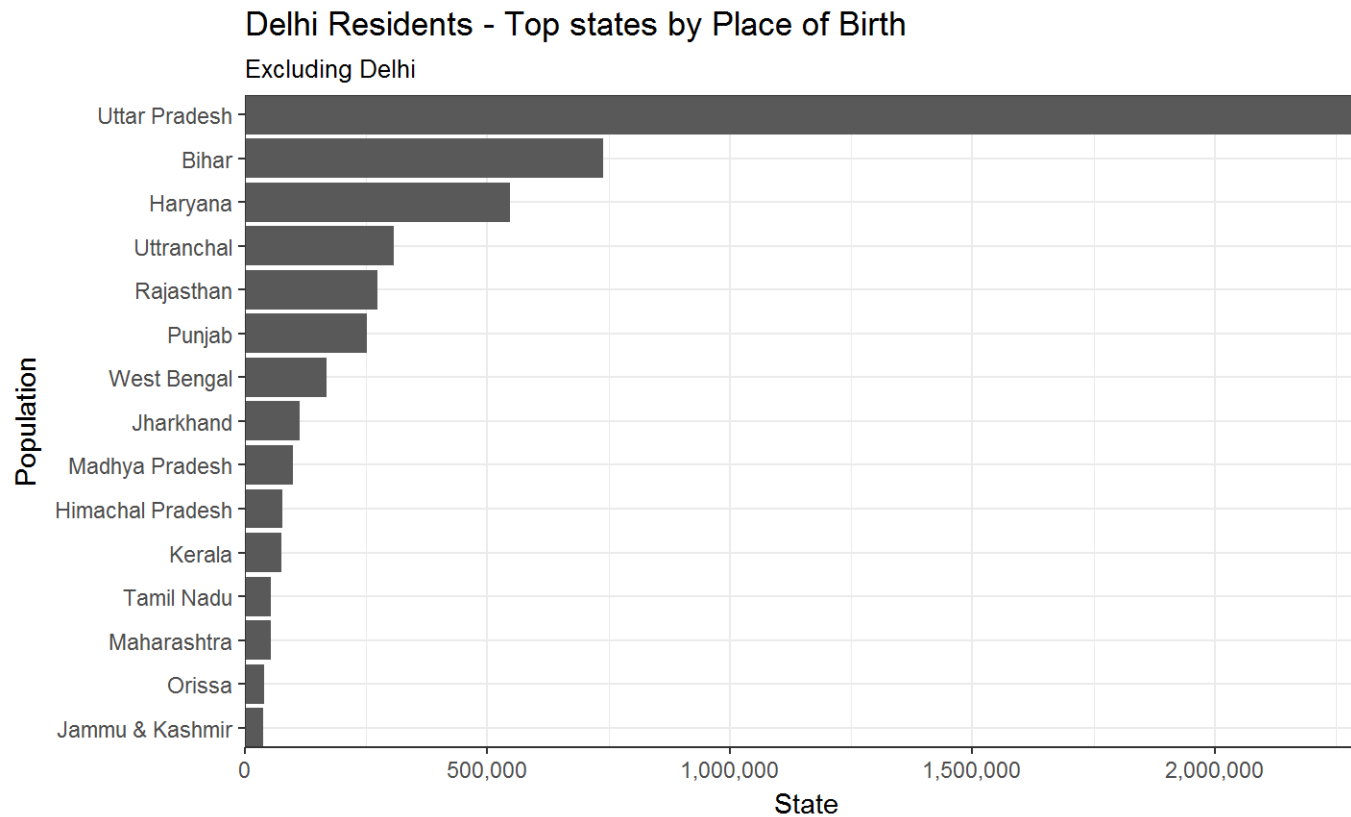
```
delhi_data_df <- readxl::read_excel('delhi.xls')
delhi_data_df %>% glimpse
```

```
## Observations: 3,960
## Variables: 15
## $ Table Name      <chr> "D0201A", "D0201A", "D020...
## $ State Code      <chr> "07", "07", "07", "07", "...
## $ District Code   <chr> "00", "00", "00", "00", "...
## $ Area Name       <chr> "UNION TERRITORY - DELHI ...
## $ Age-Group       <chr> "All ages", "All ages", "...
## $ Birth Place     <chr> "Total Population", "Born...
## $ Place of Enumeration - Total Persons <dbl> 13850507, 13522592, 82042...
## $ Place of Enumeration - Total Males   <dbl> 7607234, 7429601, 4445651...
## $ Place of Enumeration - Total Females <dbl> 6243273, 6092991, 3758579...
## $ Place of Enumeration - Rural Persons <dbl> 944727, 936324, 555531, 4...
## $ Place of Enumeration - Rural Males   <dbl> 522087, 517211, 324676, 3...
## $ Place of Enumeration - Rural Females <dbl> 422640, 419113, 230855, 1...
## $ Place of Enumeration - Urban Persons <dbl> 12905780, 12586268, 76486...
## $ Place of Enumeration - Urban Males   <dbl> 7085147, 6912390, 4120975...
## $ Place of Enumeration - Urban Females <dbl> 5820633, 5673878, 3527724...
```

Exploring the dataset

```
library(ggplot2)
gg <- delhi_data_df %>%
  filter(`Age-Group` %in% 'All ages',
         `Area Name` %in% 'UNION TERRITORY - DELHI 07') %>%
  select(`Birth Place`, `Place of Enumeration - Total Persons`) %>%
  setNames(c('var', 'val')) %>%
  slice(8:42) %>%
  arrange(desc(val)) %>%
  head(n = 15) %>%
  mutate(var = factor(var, levels = rev(var))) %>%
  ggplot() +
  geom_bar(aes(x = var, y = val), stat = 'identity') +
  coord_flip(expand = FALSE) +
  theme_bw() +
  xlab('Population') +
  ylab('State') +
  scale_y_continuous(labels = scales::comma) +
  ggtitle('Delhi Residents - Top states by Place of Birth', subtitle = 'Excluding Delhi')
```

Exploring the dataset



Go play!

Download data for other states from this catalog, and explore inter-state migration patterns.

Take a look at the code in <https://github.com/sainathadapa/population-pyramid-states-india>, in case you are stuck.

Thanks!

- If you are facing problems with the functions, or have ideas for contribution, please create issues on the Github page
- You can contact me via twitter @sainathadapa
- This presentation was written in RMarkdown. Code for this presentation is available at <https://gist.github.com/sainathadapa/72a2412f512dde220307b1f907dc62f6>