

# **CSCI 599**

# **Content Detection and Analysis for Big Data**

**Teague Ashcraft**  
**Mohamad El Sayed**  
**Khanh Duy Le**  
**Kevin Le**

# Three Data Sets

- **Forest Fires Data Set**



- **Criminal Recidivism Data Set**

Definition: Describes an individual's relapse into criminal behavior/jail

- **Zip Code Data Set**



# Forest Fires Data Set

**LOW VOLUME**

**LOW VARIETY**

**LOW VELOCITY**

**HIGH VERACITY**

**HIGH VALUE**

# Criminal Recidivism Data Set

**LOW VOLUME**

**LOW VARIETY**

**LOW VELOCITY**

**HIGH VERACITY**

**HIGH VALUE**

# Income / Tax Zip Code Data Set

Gross income, returns, wages, salaries, dividends, interests, etc.

**LOW VOLUME**

**LOW VELOCITY**

**HIGH VALUE**

The screenshot shows the IRS website's page for SOI Tax Stats - Individual Income Tax Statistics - 2015 ZIP Code Data (SOI). The page has a dark blue header with the IRS logo, a search bar, and navigation links for Help, News, Language, Charities & Nonprofits, and Tax Pros. Below the header is a navigation bar with tabs for File, Pay, Refunds, Credits & Deductions, and Forms & Instructions. The main content area has a breadcrumb trail: Home > Our Agency > Tax Statistics > SOI Tax Stats Individual Income Tax Statistics 2015 ZIP Code Data SOI. The title is "SOI Tax Stats - Individual Income Tax Statistics - 2015 ZIP Code Data (SOI)". On the left is a sidebar with a "Volunteer" section and a "Tax Statistics" section containing links for Taxpayer Compliance, Products and Publications, Individual Tax, Business Tax, By Form, Charitable, Estate and Gift, and IRS Data Book. The main content area has a "Return to Tax Stats home page" link, a "ZIP Code Data - 2015" section with a description of the data files, a "Documentation" section with a link to the ZIP Code Data Users Guide and Record Layouts, a "ZIP Code Data" section with links to .csv Files (2015) and Zip File (2015), and a "ZIP Code Data by Individual State" section with a map of the United States.

IRS

Search

Help News Language | Charities & Nonprofits Tax Pros

File Pay Refunds Credits & Deductions Forms & Instructions

Home > Our Agency > Tax Statistics > SOI Tax Stats Individual Income Tax Statistics 2015 ZIP Code Data SOI

**SOI Tax Stats - Individual Income Tax Statistics - 2015 ZIP Code Data (SOI)**

English

Volunteer

Tax Statistics

Taxpayer Compliance

Products and Publications

Individual Tax

Business Tax

By Form

Charitable

Estate and Gift

IRS Data Book

Procurement

Privacy Policy

Freedom of Information Act

Return to [Tax Stats](#) home page

**ZIP Code Data - 2015**

Data files are available for download in Comma Separated Values files (.csv file extension) where data items are separated by commas and each new line represents a new row, and Excel files by State (.xls file extension).

**Documentation**

[ZIP Code Data Users Guide and Record Layouts](#) (Word Document)

**ZIP Code Data**

**.csv Files**

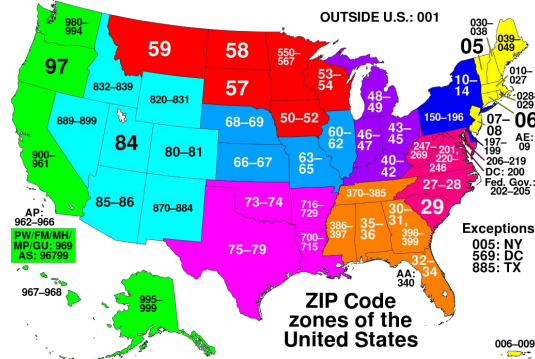
[2015](#) (all States, includes AGI) [2015](#) (all States, does not include AGI)

**Zip File** (contains the Documentation, the .csv files, and all State files in Excel format)

[2015](#)

**ZIP Code Data by Individual State**

WA MT ND VT ME



**LOW VARIETY**

**HIGH VERACITY**

# Concluding Remarks

Income Tax/Zip Code Data Set

Criminal Recidivism Data Set

Forest Fires Data Set

- What are the unintended consequences of these data sets?
  - Theft
  - Targeted advertising
- Can they be combined?
  - Income Tax/Zip Code + Criminal Recidivism
  - Income Tax/Zip Code + Forest Fires
  - Criminal Recidivism + Forest Fires