China's Corruption Investigations Dataset (CCID)
2011—2016

Yuhua Wang
yuhuawang@fas.harvard.edu
https://scholar.harvard.edu/yuhuawang/home
May 11, 2020

Description

The China's Corruption Investigations Dataset includes information on almost 20,000 officials who were investigated during Xi Jinping's anti-corruption campaign.

I collected the data from Tencent—the largest Internet company in China. During the Xi Jinping's anti-corruption campaign, Tencent launched a searchable online database of all corruption investigations across China since 2011.[1] Based on information provided by Party disciplinary committees, courts, and procuratorates from the central to local levels, Tencent's database includes each official's name, position, locality, rank, and reason for investigation. In August 2016, I used Python to scrape Tencent's website and organized the database in an analysis-ready format. To verify this database and ensure that every investigation was made public, I ran an Internet search on every name to find its original source and record the date of the announcement.

The Tencent database has two advantages. First, it is the most comprehensive, public database on China's corruption investigations. It synthesizes information from official statistics at all levels of government and from all branches. Second, Tencent has provided this online database for Internet users to search how many officials in their hometowns have been investigated for corruption. By clicking on their hometowns on a drop-down list, this online interface reports the total number of investigations. This is the only place Chinese citizens can find out this number in a single click, and the database is widely circulated via Tencent's app—WeChat—China's most popular social network app, which has over 800 million users.

How to cite the data

The database should be referred to in publications as "China's Corruption Investigations Dataset" with a link to the Harvard Dataverse url. Users are free to publish the results of their analyses without restrictions, with the understanding that the source of the data will be credited as the following:

Wang, Yuhua and Bruce Dickson. 2020. "How Corruption Investigations Undermine Regime Support: Evidence from China." Working Paper. Available at SSRN: https://ssrn.com/abstract=3086286.

[1] See http://news.qq.com/zt2016/fanfu_ccdi/index.htm (accessed February 7, 2017).

The P.I. would greatly appreciate receiving copies of any papers that employ these data, whether published or unpublished (yuhuawang@fas.harvard.edu).

Codebook

caseid
Unique identification number for each investigated official

province
2-digit province number

prefecture
Prefecture name in Chinese

county
County name in Chinese

province_id
6-digit province number

prefecture_id
6-digit prefecture number

county_id
6-digit county number

name
Official's name in Chinese

position
Official's position in Chinese

reason
Reason for investigation

rank
Official's rank (1=state level (正国级) to 10=deputy office level (副科级))

Year
Year of the announcement of the investigation

Month
Month of the announcement of the investigation