Codebook: Chinese provincial-level protest data

2021-05-26 18:42:14

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	720
Number of variables	11

Codebook summary table

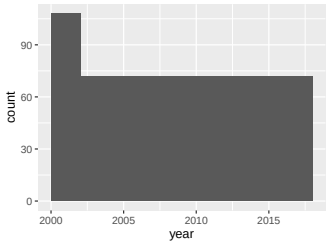
Label	Variable	Class	# unique values	Missing	Description
	year	numeric	20	5.00 %	
	region	character	32	18.19 %	
	nontaxrev	numeric	372	48.33 %	
	subin	numeric	588	18.33 %	
	subsidies	numeric	8	65.00 %	
	shared_income	numeric	591	16.11 %	
	otherincome	numeric	126	80.69 %	
	totincome	numeric	125	82.78 %	
	T_goods	numeric	560	17.50 %	
	N_SOE	numeric	215	64.86 %	
	n_protest	numeric	67	62.50 %	

Variable list

year

Year of Observation

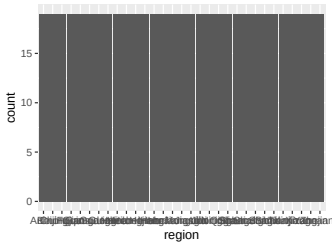
Feature	Result
Variable type	numeric
Number of missing obs.	36 (5 %)
Number of unique values	19
Median	2009
1st and 3rd quartiles	2004; 2014
Min. and max.	2000; 2018



region

Region of Observation

Feature	Result
Variable type	character
Number of missing obs.	131 (18.19 %)
Number of unique values	31
Mode	“Anhui”



- Observed factor levels: ”Anhui”, ”Beijing”, ”Chongqin”, ”Fujian”, ”Gansu”, ”Guangdong”, ”Guangxi”, ”Guizhou”, ”Hainan”, ”Hebei”, ”Heilongjiang”, ”Henan”, ”Hubei”, ”Hunan”, ”Inner Mongolia”, ”Jiangsu”, ”Jiangxi”, ”Jilin”, ”Liaoning”, ”Ningxia”, ”Qinghai”, ”Shaanxi”, ”Shangdong”, ”Shanghai”, ”Shanxi”, ”Sichuan”, ”Tianjin”, ”Xinjiang”, ”Xizang”, ”Yunnan”, ”Zhejiang”.

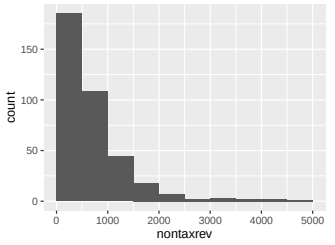
---

nontaxrev

Non-Tax Revenue (100 million yuan)

Non-Tax Revenue reported in the Statistical Yearbook of China. (In Chinese: 各地区财政收入-非税收入)

Feature	Result
Variable type	numeric
Number of missing obs.	348 (48.33 %)
Number of unique values	371
Median	503.24
1st and 3rd quartiles	237.24; 862.57
Min. and max.	8.93; 4526.99



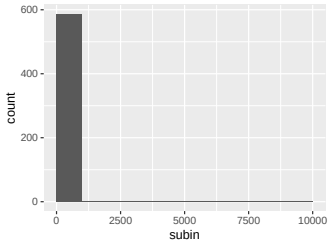
---

subin

Subsidies Income (100 million yuan)

Subsidies Income reported in the Statistical Yearbook of China. (In Chinese - 各地区财政收入- 专项收入).

Feature	Result
Variable type	numeric
Number of missing obs.	132 (18.33 %)
Number of unique values	587
Median	106.56
1st and 3rd quartiles	52.32; 216.19
Min. and max.	0.2; 9630



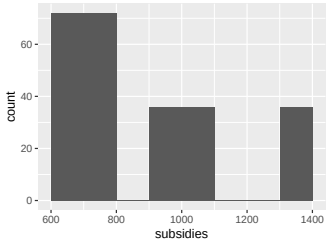
---

subsidies

Sum of special price subsidies (100 million yuan)

In particular years, the Statistical Yearbook of China reported amounts of specific subsidies received by the provinces. For example, State-Owned Enterprise Capital Loss Subsidies (国有企业计划亏损补贴), Grain, Cotton and Gasoline Price subsidies (粮棉油价格补贴), Consumer Goods Price Subsidies (平抑物价等补贴), Meat Price Subsidies (肉食品价格补贴) and other price subsidies. These are temporary subsidies that are only granted once the national average price level has reached a required level. For details, see http://www.gov.cn/zhengce/content/2018-04/13/content_5281683.htm.

Feature	Result
Variable type	numeric
Number of missing obs.	468 (65 %)
Number of unique values	7
Median	795.8
1st and 3rd quartiles	645.07; 1042.28
Min. and max.	617.28; 1387.52

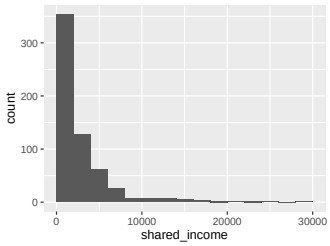


shared_income

Shared Income (100 million yuan)

Data from the Statistical Year Book of China. Shared Income is calculated as the sum of national tax income distributed for each province. Due to a tax-sharing regime in China, provinces will receive a small percentage of the tax income collected at the national level. This shared income variable is thus a measure of the sum of national tax incomes redistributed to the provincial level.

Feature	Result
Variable type	numeric
Number of missing obs.	116 (16.11 %)
Number of unique values	590
Median	1648.06
1st and 3rd quartiles	838.95; 3313.89
Min. and max.	0; 28647.89

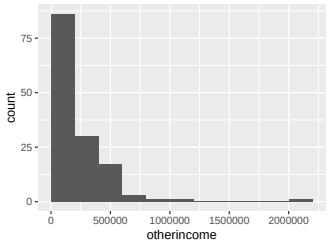


otherincome

Other Income

Other Income reported in the statistical year book of China. (In Chinese - 其他收入).

Feature	Result
Variable type	numeric
Number of missing obs.	581 (80.69 %)
Number of unique values	125
Median	140263
1st and 3rd quartiles	24477; 291431.5
Min. and max.	0; 2018731

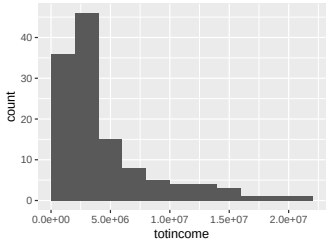


totincome

Total Income

A sum of all incomes for each province.

Feature	Result
Variable type	numeric
Number of missing obs.	596 (82.78 %)
Number of unique values	124
Median	2995123.5
1st and 3rd quartiles	1704444; 5349053.5
Min. and max.	73082; 21794608

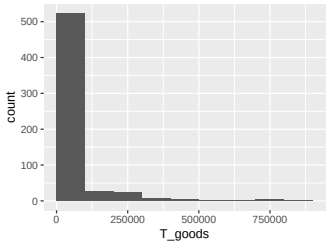


T_goods

Total of Volume of Goods Traded

Provincial-level (determined by the location of the firm) data for the total of volume of goods traded reported in the statistical year book of China. (In Chinese - 各地区进出口商品总值 (按经营单位所在地分))

Feature	Result
Variable type	numeric
Number of missing obs.	126 (17.5 %)
Number of unique values	559
Median	8709.54
1st and 3rd quartiles	2031.3; 32888.58
Min. and max.	0; 873772.66

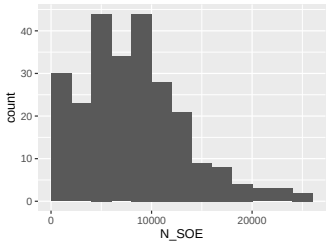


N_SOE

Number of State-Owned Enterprises

Number of State-Owned Enterprises reported by the Statistical Year Book of China. (In Chinese - 按地区和控股情况分企业法人单位数 (2017 年))

Feature	Result
Variable type	numeric
Number of missing obs.	467 (64.86 %)
Number of unique values	214
Median	7954
1st and 3rd quartiles	4951; 11102
Min. and max.	0; 24617



n_protest

Number of Protests identified by the GDELT 2.0 Events Database.

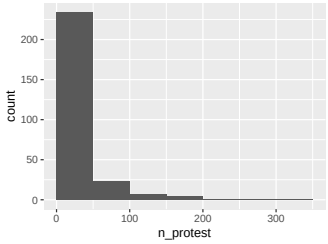
The protest events data is acquired by using the following query to the GDELT 2.0 Events Database:

```
SELECT COUNT(*) as n_events, LEFT(CAST(SQLDATE as STRING),4) as year
FROM  `gdelt-bq.gdeltv2.events`
WHERE SQLDATE>19970901 and SQLDATE < 20200901 and EventCode = "PROTEST" and ActionGeo_CountryCode ="CN"
GROUP BY year
ORDER BY year;
```

Number of Protests identified by the GDELT 2.0 Events Database. Since the events recorded by GDELT is geocoded, the provincial location of the event is determined by reverse geo-coding the data.

An example of the reverse geocoding program can be found in `reverse-geocoding.R`.

Feature	Result
Variable type	numeric
Number of missing obs.	450 (62.5 %)
Number of unique values	66
Median	7
1st and 3rd quartiles	3; 21
Min. and max.	1; 318



Report generation information:

- Created by: Hei Yin Kyle Chan
- Report creation time: Wed May 26 2021 18:42:14
- Report was run from directory: `C:/Users/Kyle/Documents/GitHub/hkmocn_paper`
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 4.0.5)]
- R version 4.0.2 (2020-06-22).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 19042)).
- Function call: `dataMaid::makeDataReport(data = china, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook_china.Rmd", checks = list(character = "showAllFactorLeve factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for china")`