Data Appendix

Eicher, “Right-to-work and Workplace Safety”

This data appendix provides definitions, coding, and summary statistics for variables in the analysis data file, *ITA_state.dta*.

This analysis dataset contains 17,276 observations, each representing one establishment in one year in OSHA’s Injury Tracking Application (ITA) database.

The distribution of observations across states is as follows:

state	Freq.	Percent	Cum.
KY	4,341	25.13	25.13
OH	12,935	74.87	100.00
Total	17,276	100.00	

In addition to *state*, there are six categorical variables which are useful only for identifying observations and have no role in the analysis at this stage: *company_name, establishment_name, street_address, city, zip_code, establishment_id*.

There are 20 variables which are relevant to the analysis:

- 15 establishment-level variables: *log_ann_empl, log_total_hours_worked, log_total_dafw_days, log_total_injuries, log_ipw, log_total_resp_cond, year, state_dummy, time_dummy_2017, state_time, naics_code, size, construction, manufacturing, private*
- Five sector-state-level variables: *empl, member, covered, pctmem100, pctcov100*

ESTABLISHMENT-LEVEL VARIABLES

Variable name: *log_ann_empl*

Source: ITA dataset - *annual_average_employees*

Values: Positive numbers

Coding: Natural log of annual average employees at a given establishment

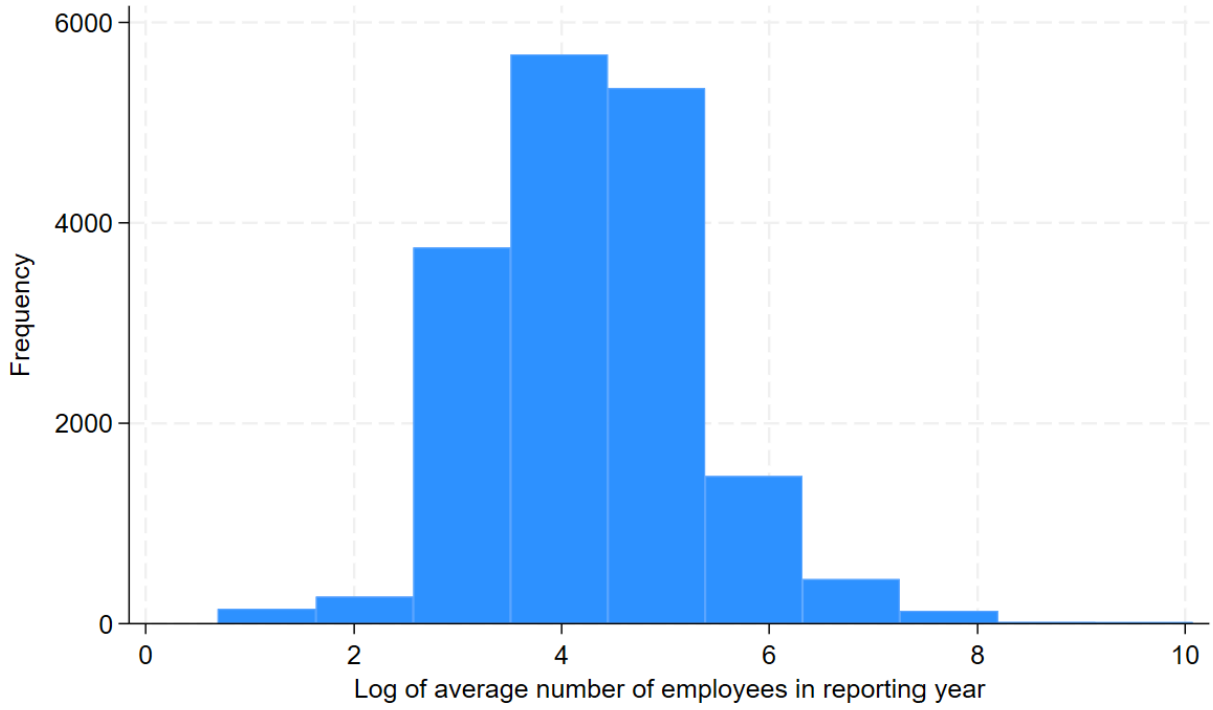
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_ann_empl	4.293812	1.040186	.6931472	10.07353	3.526361	4.248495	4.94876

Histogram:



Variable name: *log_total_hours_worked*

Source: ITA dataset - *total_hours_worked*

Values: Positive numbers

Coding: Natural log of total worker-hours for the establishment in a given year

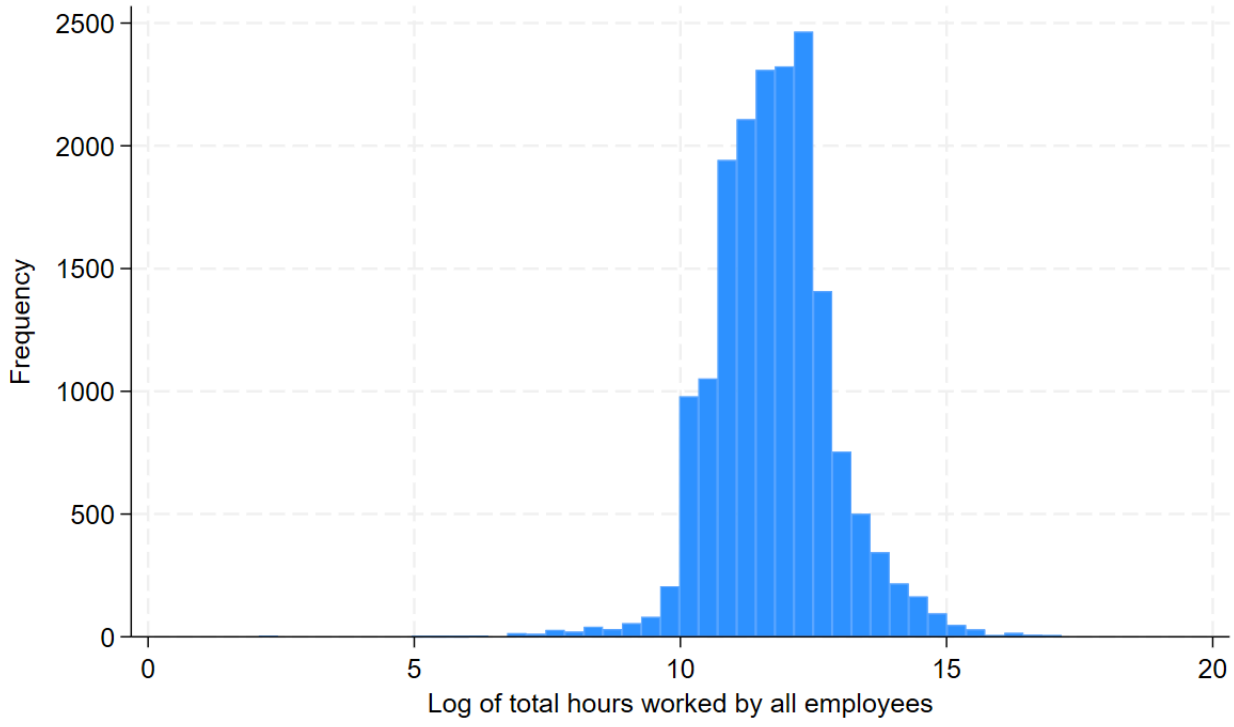
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_total~ed	11.75075	1.111814	2.079442	17.16174	11.03587	11.73665	12.37982

Histogram:



Variable name: *log_total_dafw_days*

Source: ITA dataset - *total_dafw_days*

Values: Nonnegative integers

Coding: Natural log of total number of days away from work across all workers in a given year

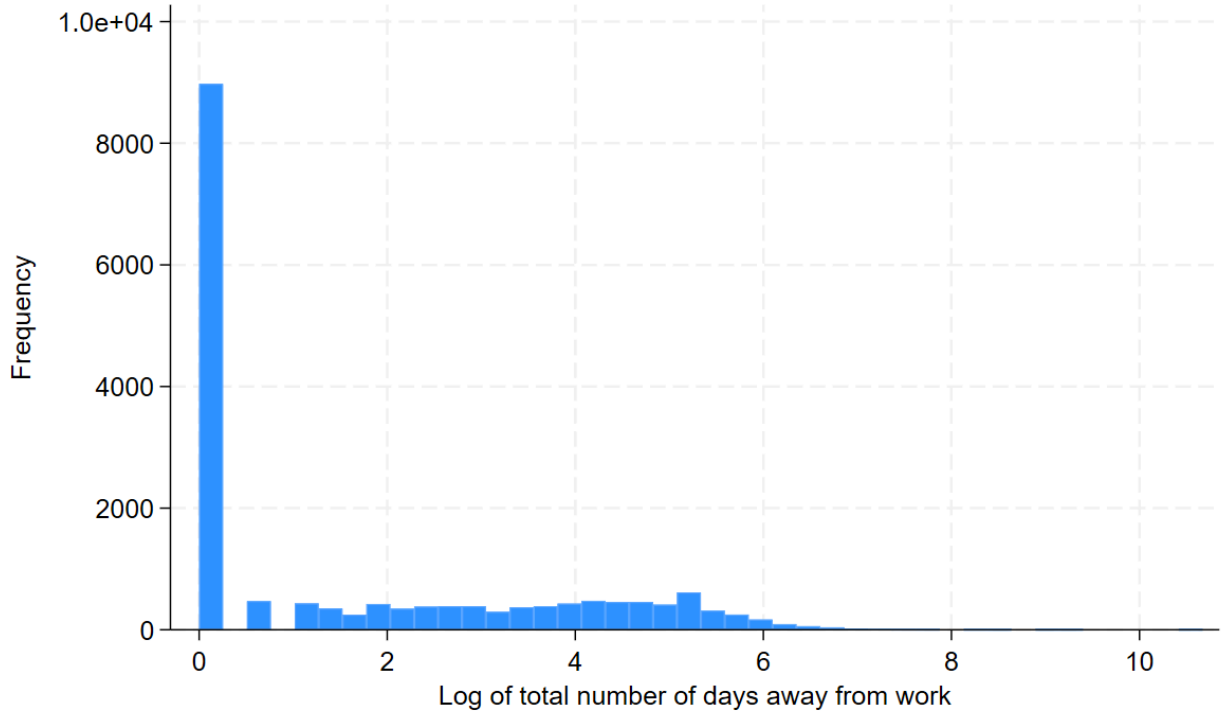
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_total~ys	1.664584	2.055326	0	10.67114	0	0	3.496508

Histogram:



Variable name: *log_total_injuries*

Source: ITA dataset - *total_injuries*

Values: Nonnegative integers

Coding: Natural log of total number of injuries at an establishment in a year

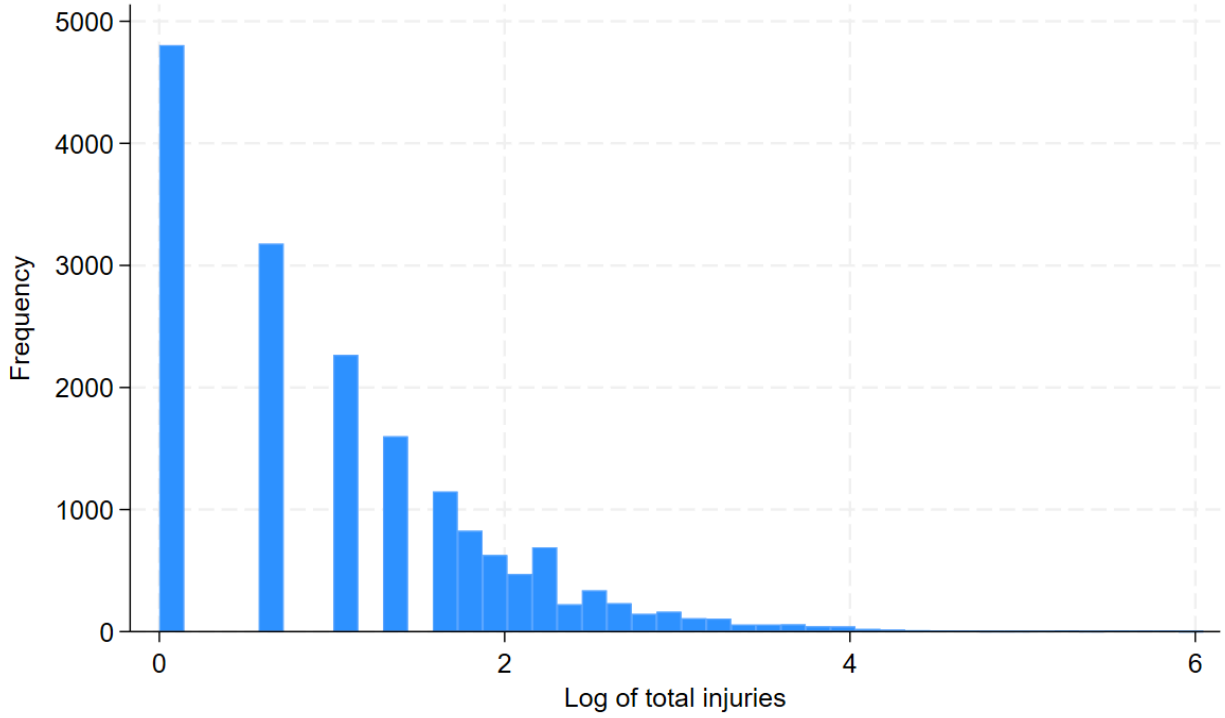
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_total~es	1.095202	.9426756	0	6.047372	0	1.098612	1.609438

Histogram:



Variable name: *log_ipw*

Source: ITA dataset - *total_injuries/annual_average_employees*

Values: Nonnegative integers

Coding: Natural log of ratio of total establishment injuries to annual average employees

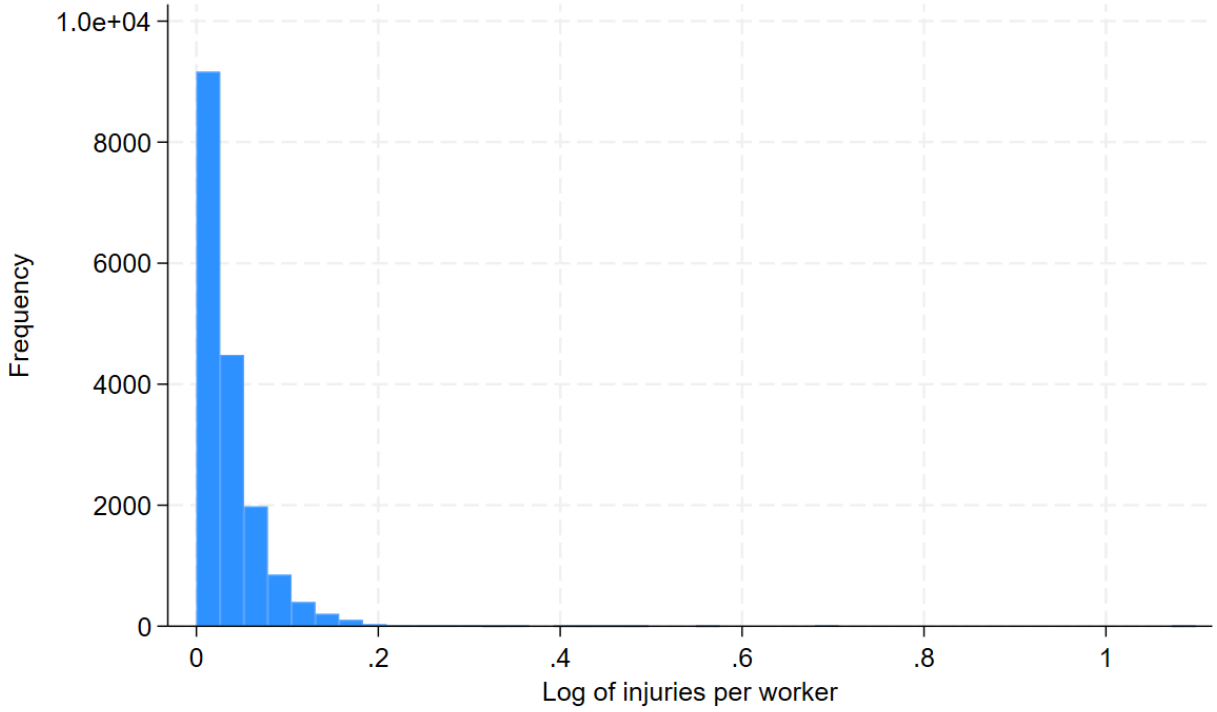
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_ipw	.032485	.040211	0	1.098612	0	.0235305	.04652

Histogram:



Variable name: *log_total_resp_cond*

Source: ITA dataset - *total_respiratory_conditions*

Values: Nonnegative integers

Coding: Total number of respiratory conditions at an establishment in a year

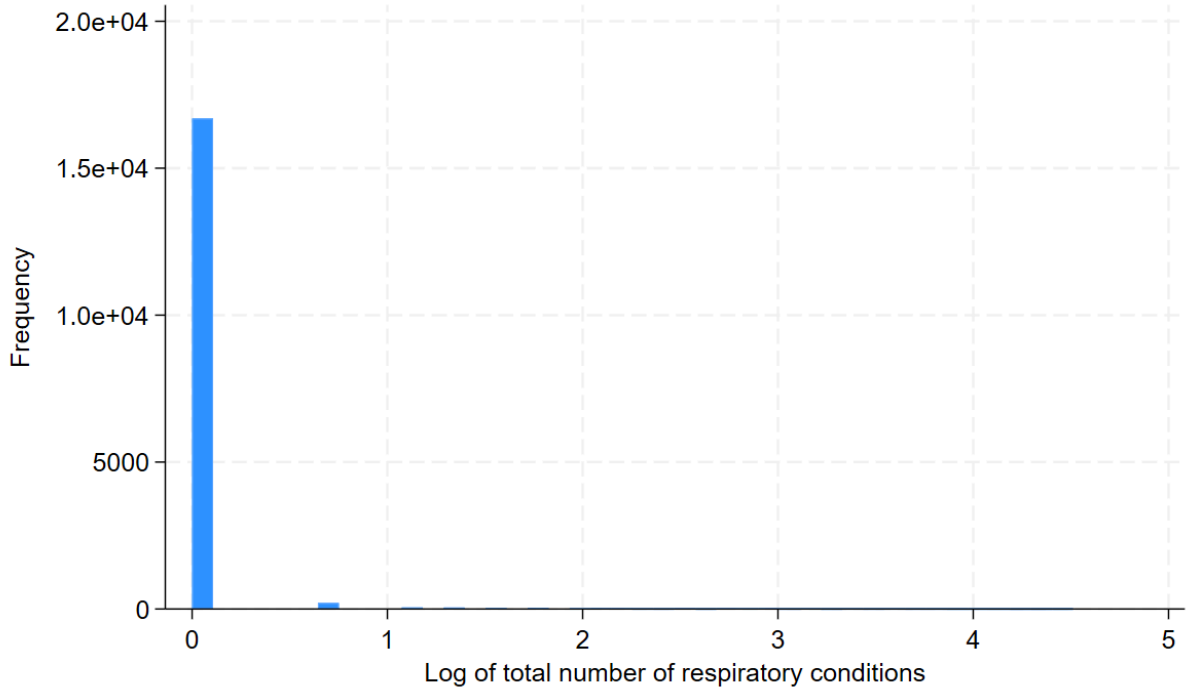
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
log_total~nd	.0575474	.3622809	0	4.510859	0	0	0

Histogram:



Variable name: *year*

Source: ITA dataset

Values: Positive integers

Coding: Year of observation

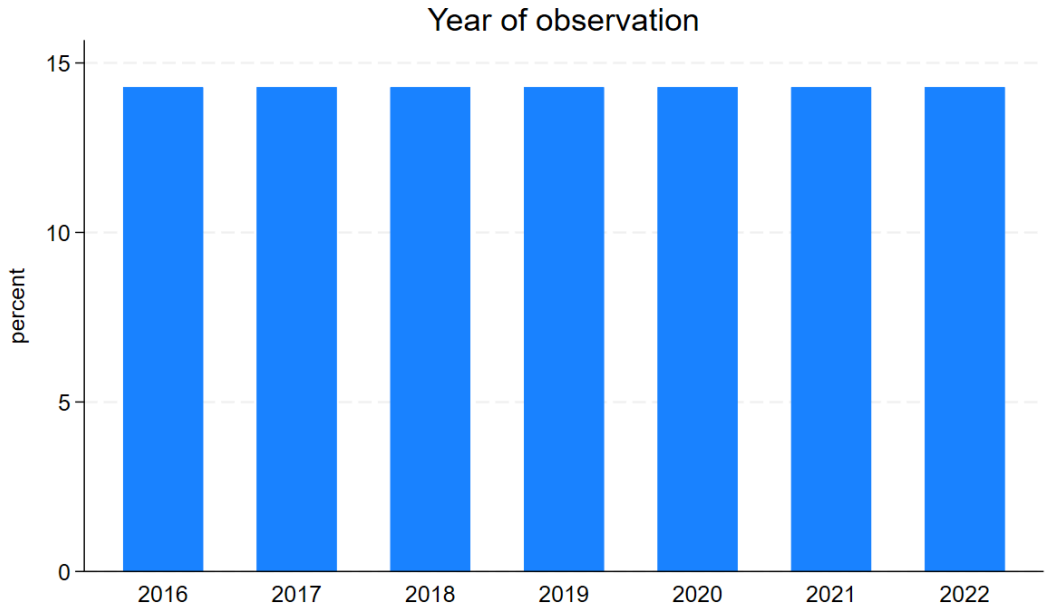
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

year	Freq.	Percent	Cum.
2016	2,468	14.29	14.29
2017	2,468	14.29	28.57
2018	2,468	14.29	42.86
2019	2,468	14.29	57.14
2020	2,468	14.29	71.43
2021	2,468	14.29	85.71
2022	2,468	14.29	100.00
Total	17,276	100.00	

Relative frequency chart:



Variable name: *state_dummy*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Dummy corresponding to state of establishment (0 = Ohio, 1 = Kentucky)

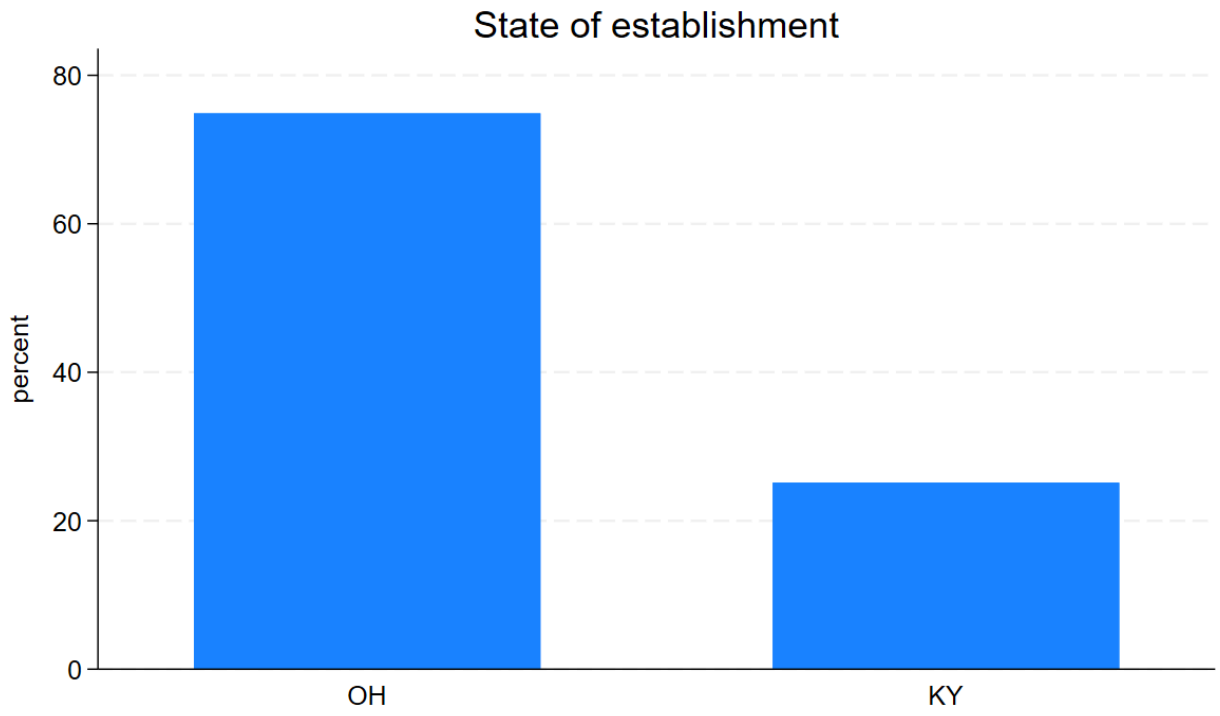
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

KY	Freq.	Percent	Cum.
0	12,935	74.87	74.87
1	4,341	25.13	100.00
Total	17,276	100.00	

Relative frequency chart:



Variable name: *time_dummy_2017*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Dummy corresponding to year of observation, relative to 2017 (0 = Pre-2017, 1 = 2017 or later)

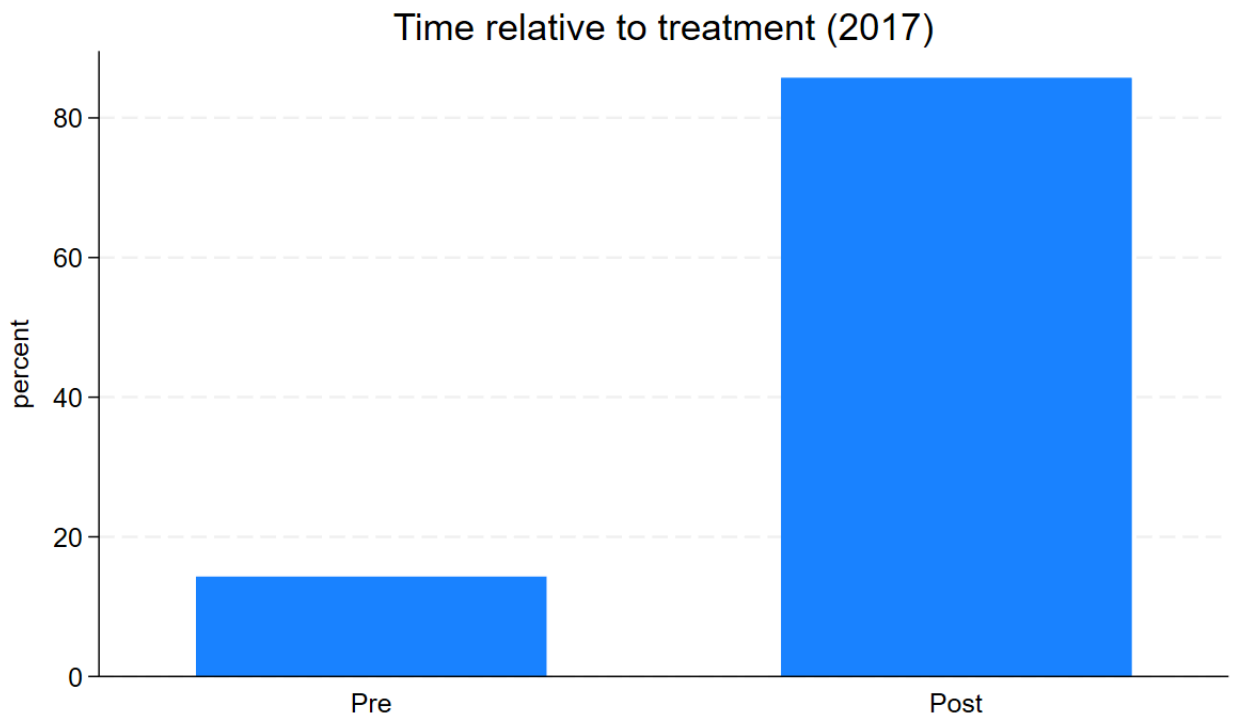
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

Post-treatment	Freq.	Percent	Cum.
0	2,468	14.29	14.29
1	14,808	85.71	100.00
Total	17,276	100.00	

Relative frequency chart:



Variable name: *state_time*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Interaction term between *state_dummy* and *time_dummy_2017* (0 = Pre-2017 OH and KY, 1 = 2017 or later KY)

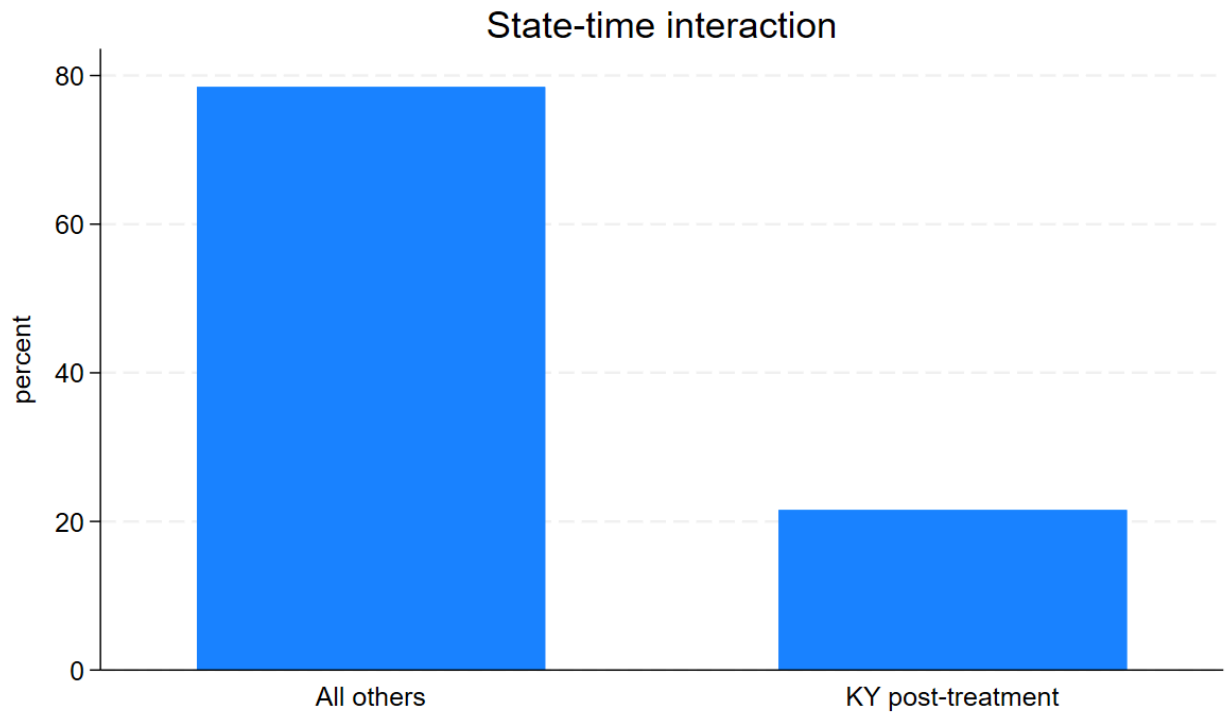
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

KY,_Post-treatment	Freq.	Percent	Cum.
0	13,555	78.46	78.46
1	3,721	21.54	100.00
Total	17,276	100.00	

Histogram:



Variable name: *naics_code1*

Source: Derived from ITA dataset

Values: Positive integer (categorical)

Coding: 2-digit NAICS code for establishment industry

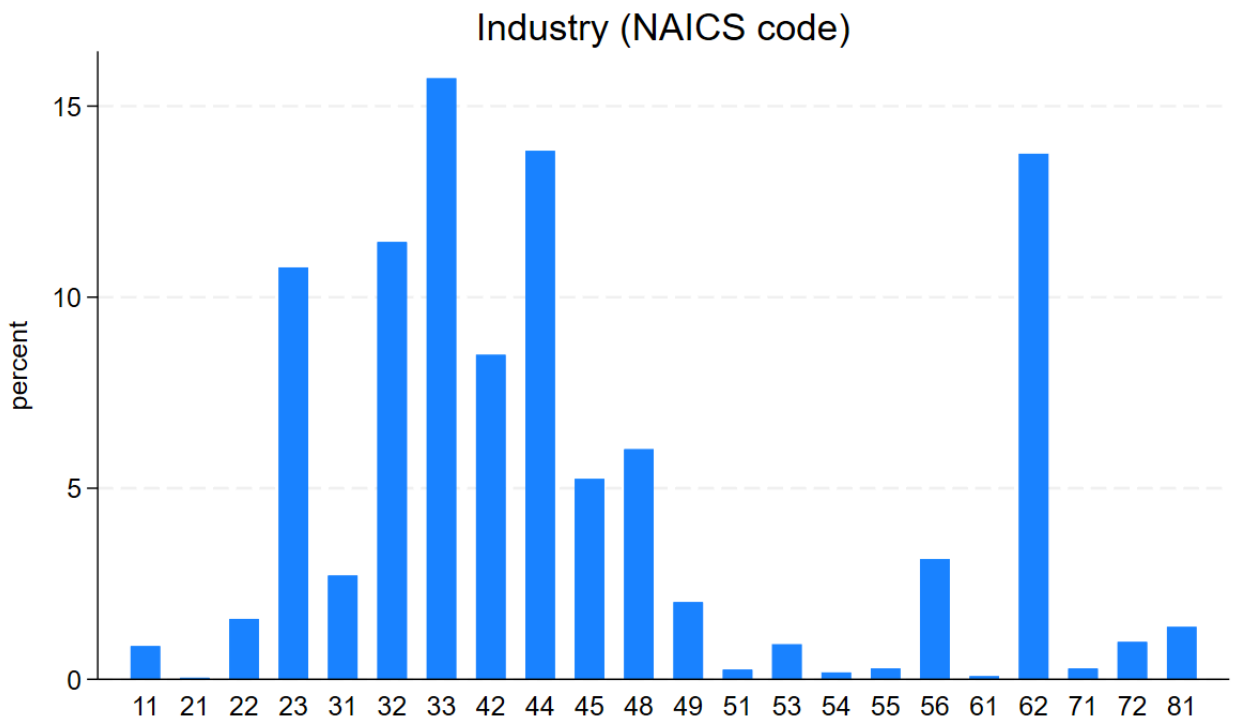
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

Part_1_of_naics_code_	Freq.	Percent	Cum.
11	150	0.87	0.87
21	7	0.04	0.91
22	273	1.58	2.49
23	1,862	10.78	13.27
31	469	2.71	15.98
32	1,977	11.44	27.43
33	2,717	15.73	43.15
42	1,467	8.49	51.64
44	2,389	13.83	65.47
45	906	5.24	70.72
48	1,040	6.02	76.74
49	349	2.02	78.76
51	44	0.25	79.01
53	158	0.91	79.93
54	31	0.18	80.11
55	49	0.28	80.39
56	543	3.14	83.53
61	14	0.08	83.61
62	2,376	13.75	97.37
71	49	0.28	97.65
72	169	0.98	98.63
81	237	1.37	100.00
Total	17,276	100.00	

Histogram:


Industry (NAICS code)
percent
0
5
10
15
11
21
22
23
31
32
33
42
44
45
48
49
51
53
54
55
56
61
62
71
72
81

Variable name: *size*

Source: ITA dataset

Values: Categorical

Coding: Dummy corresponding to an establishments number of average annual employees (1: 0-24; 2: 25-249; 3: 250+)

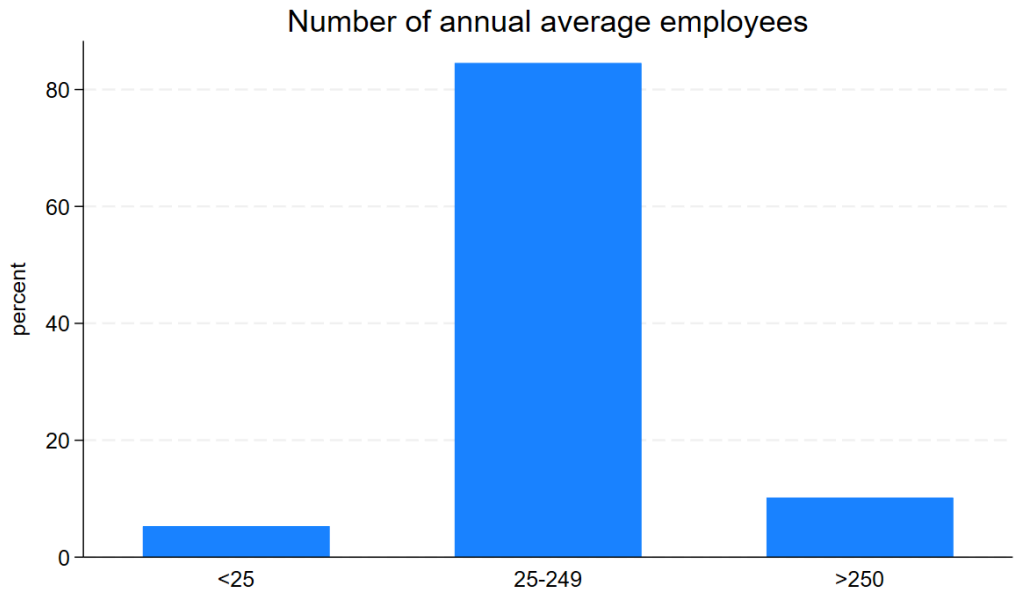
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

size	Freq.	Percent	Cum.
1	916	5.30	5.30
2	14,602	84.52	89.82
3	1,758	10.18	100.00
Total	17,276	100.00	

Relative Frequency chart:



Variable name: *construction*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Dummy corresponding to an establishment being in the construction sector (0 = No, 1 = Yes)

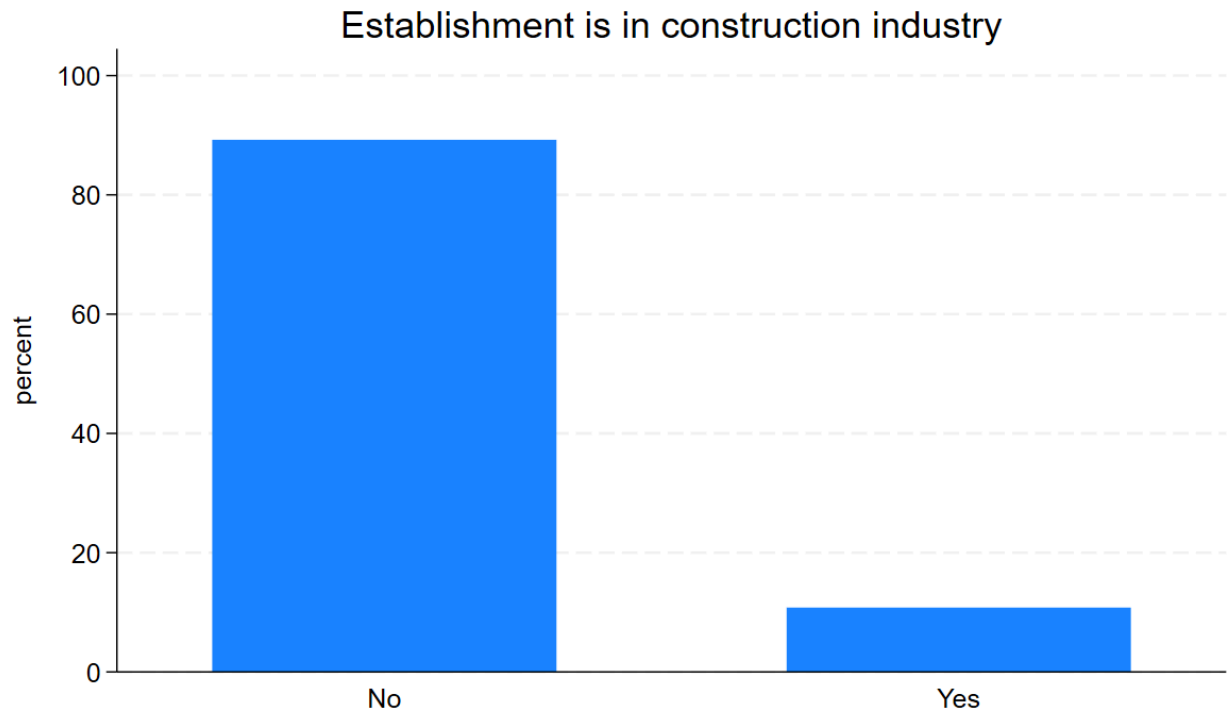
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

Constructio n	Freq.	Percent	Cum.
0	15,414	89.22	89.22
1	1,862	10.78	100.00
Total	17,276	100.00	

Relative frequency chart:



Variable name: *manufacturing*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Dummy corresponding to an establishment being in the manufacturing sector (0 = No, 1 = Yes)

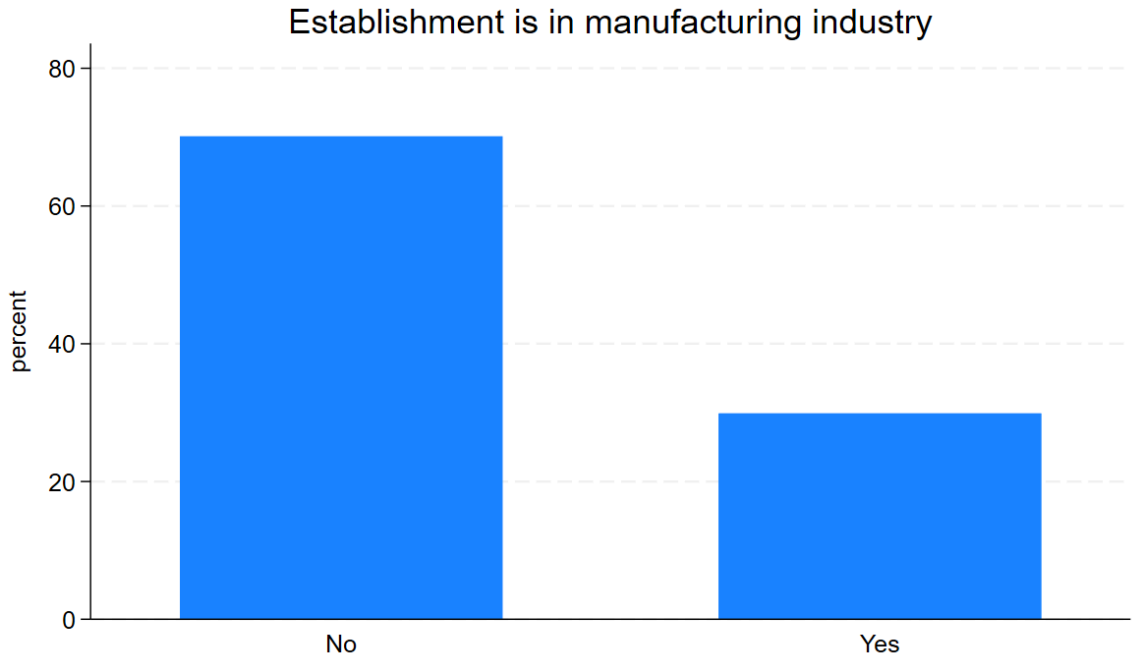
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

Manufacturing	Freq.	Percent	Cum.
0	12,113	70.11	70.11
1	5,163	29.89	100.00
Total	17,276	100.00	

Relative frequency chart:



Variable name: *private*

Source: Derived from ITA dataset

Values: Binary (0 or 1)

Coding: Dummy corresponding to an establishment being in the private sector, outside of construction and manufacturing (0 = No, 1 = Yes)

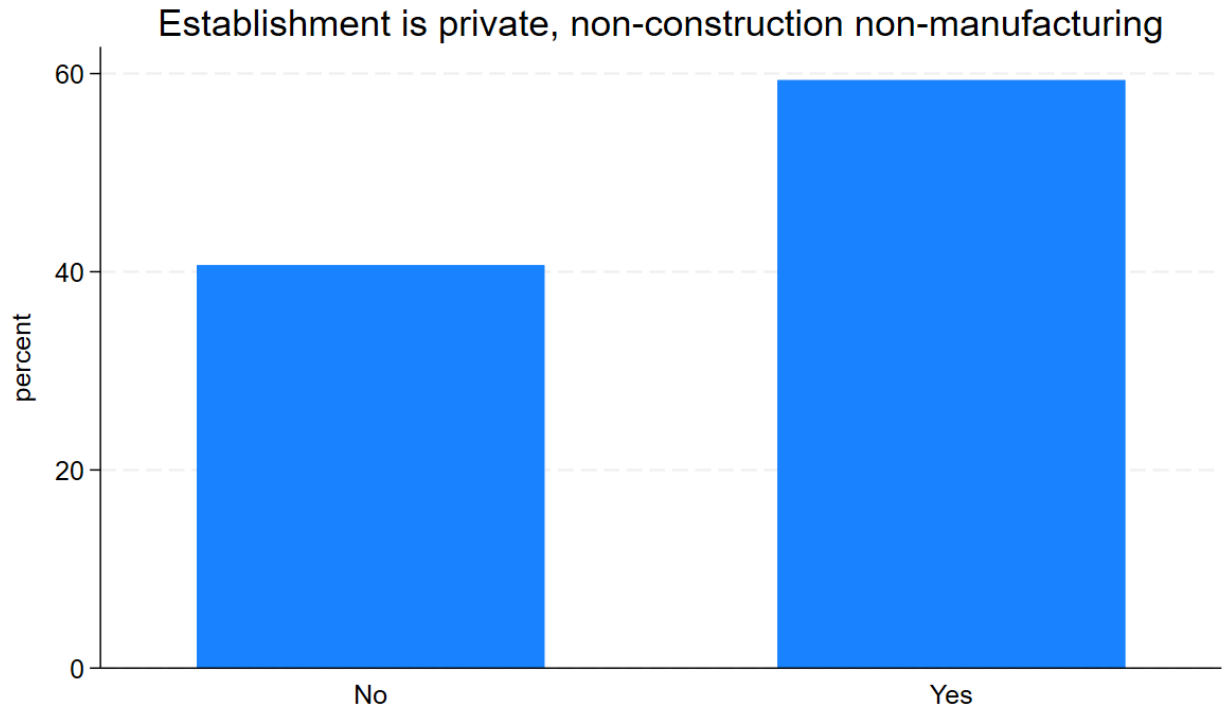
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Frequency table:

Private	Freq.	Percent	Cum.
0	7,025	40.66	40.66
1	10,251	59.34	100.00
Total	17,276	100.00	

Relative frequency chart:



STATE-SECTOR VARIABLES

Variable name: *empl*

Source: Hirsch, Macpherson, Even (2023)

Values: Positive real numbers

Coding: Number of employed persons in a given sector and state for a given year

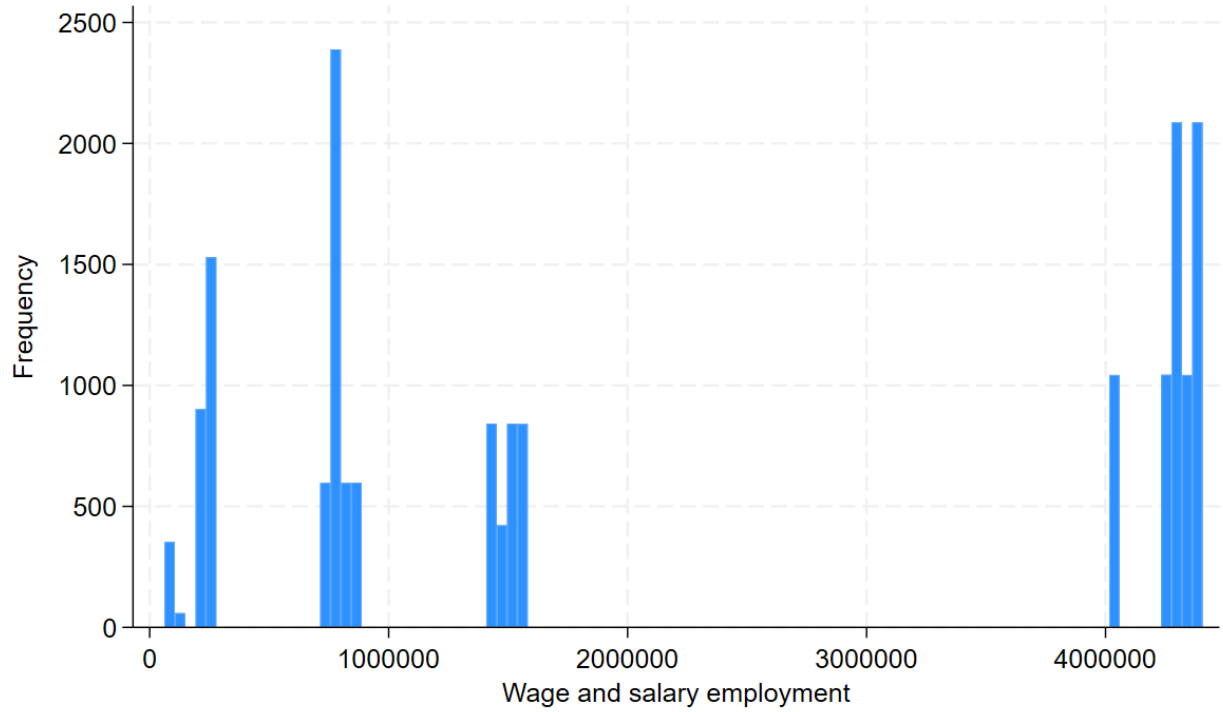
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
empl	2297612	1749739	61702.84	4406358	779922.6	1517434	4282457

Histogram:



Variable name: *member*

Source: Hirsch, Macpherson, Even (2023)

Values: Positive real numbers

Coding: Number of employed persons who are members of a union in a given sector and state for a given year

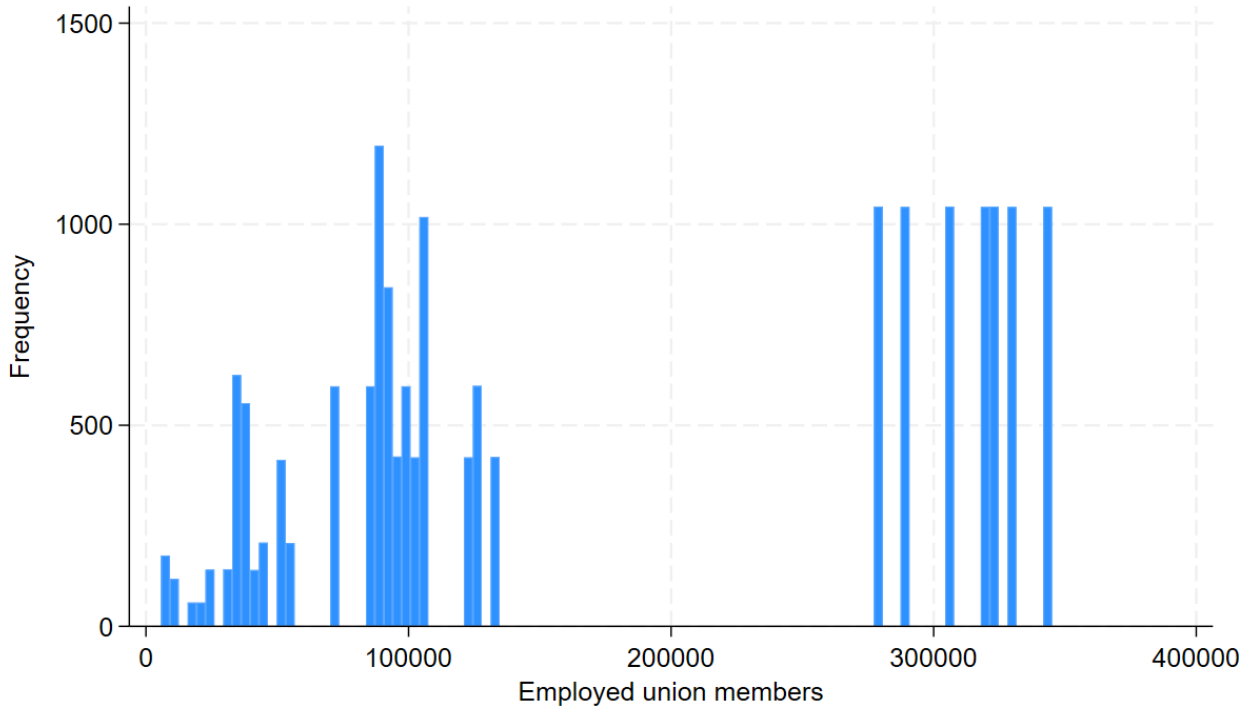
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
member	179126.4	117782.6	5695.203	345099.8	87433.22	123227.4	305339.8

Histogram:



Variable name: *covered*

Source: Hirsch, Macpherson, Even (2023)

Values: Positive real numbers

Coding: Number of employed persons covered by a CBA in a given sector and state for a given year

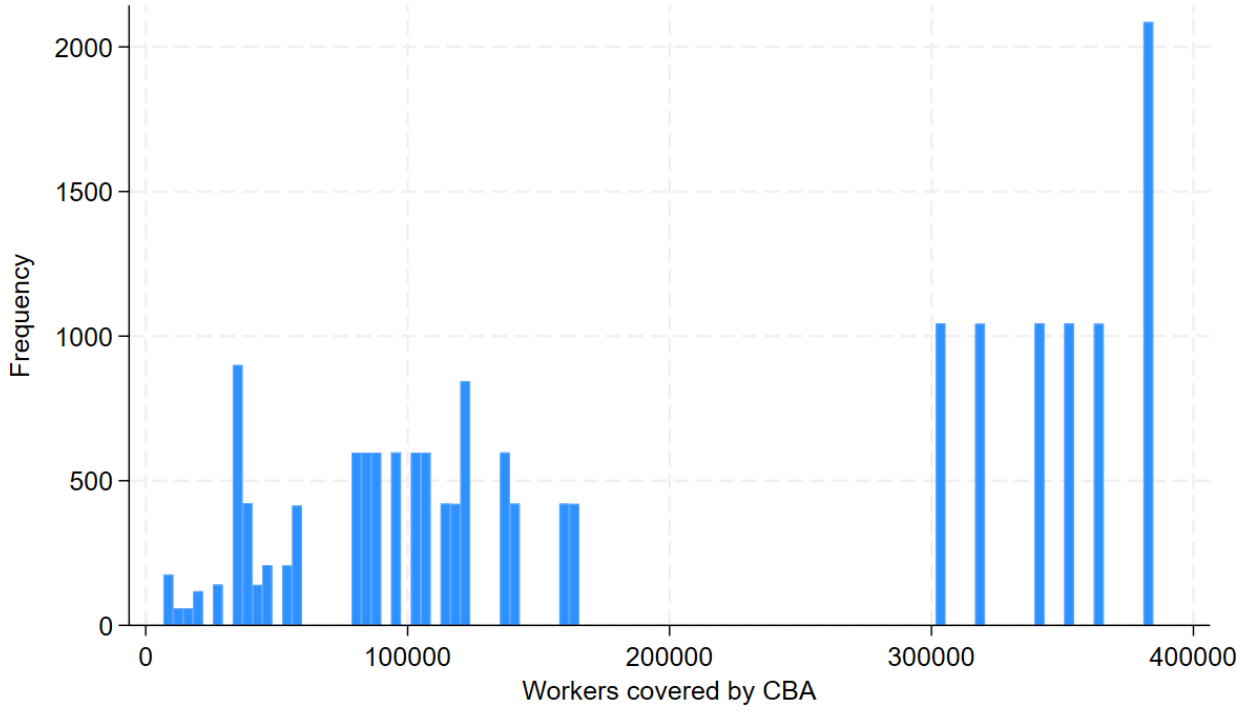
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
covered	201142.2	132134.7	6744.754	384680.5	87808.42	137124.8	342298.3

Histogram:



Variable name: *pctmem100*

Source: Hirsch, Macpherson, Even (2023)

Values: Positive real numbers

Coding: Percent of employed persons who are members of a union in a given sector and state for a given year

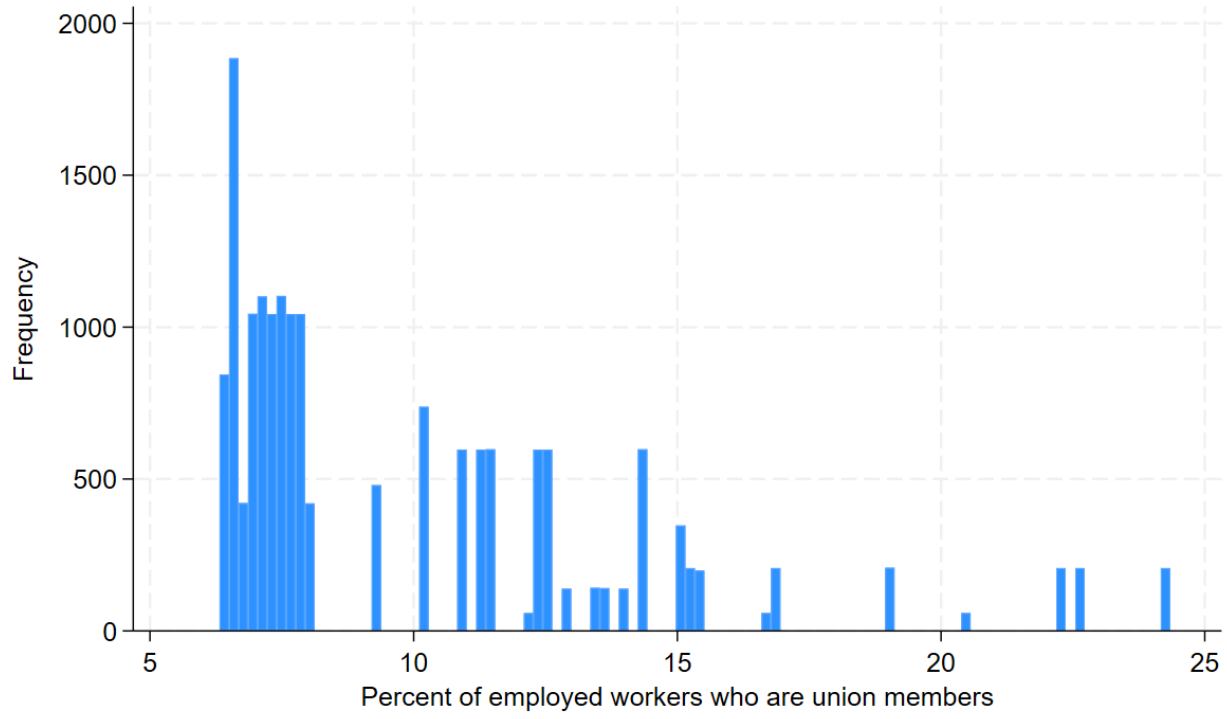
Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
pctmem100	9.856964	3.98664	6.318544	24.34675	7.103959	7.854594	11.83307

Histogram:



Variable name: *pctcov100*

Source: Hirsch, Macpherson, Even (2023)

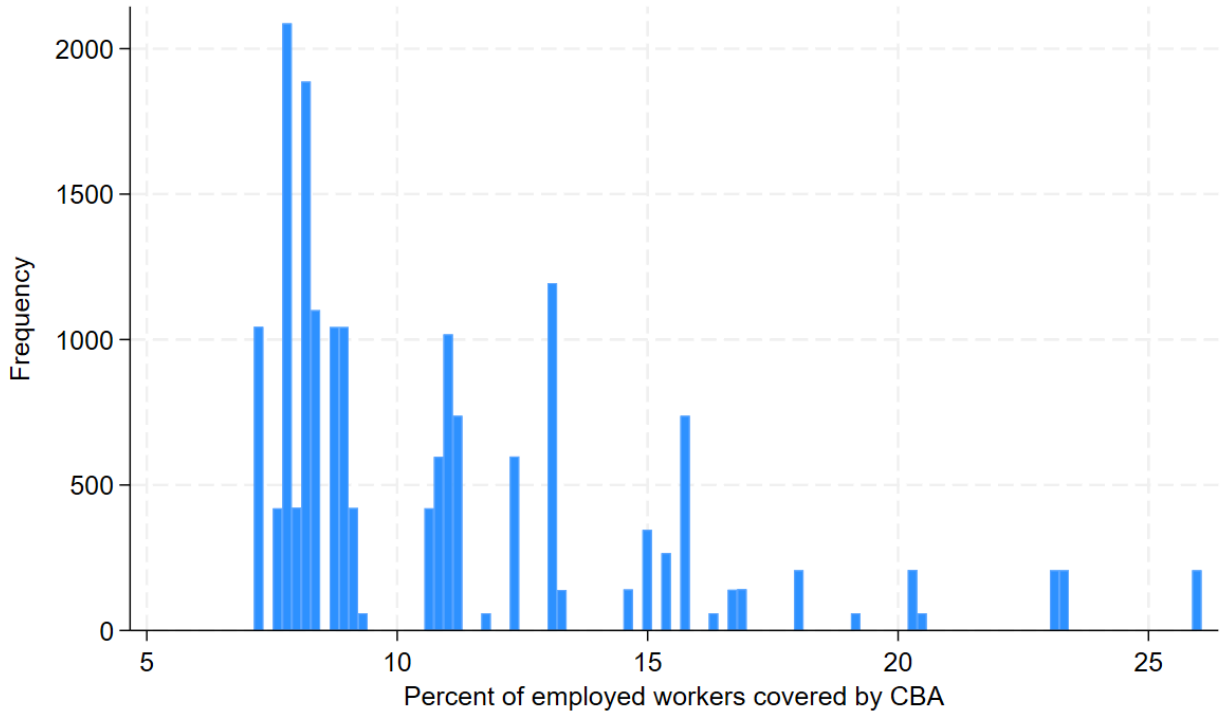
Values: Positive real numbers

Coding: Percent of employed persons covered by a CBA in a given sector and state for a given year

Number of missing observations: 0

Total number of observations (non-missing and missing): 17,276

Summary statistics:

Variable	Mean	SD	Min	Max	p25	p50	p75
pctcov100	10.87846	3.976241	7.137051	26.05936	8.19733	8.927377	12.73601

Histogram: