

紙の統計表をテキストデータ化する ネ申Excelのルーツ・芸術・情報抽出

2021年12月3日@2021年度「NDLデジタルライブラリーカフェ」
有本 寛(一橋大学)
arimotoy@ier.hit-u.ac.jp

自己紹介

- 名前: 有本 寛 (一橋大学経済研究所)
- 仕事: 経済学者 (開発経済学, 経済史, 農業経済学)
 - 現地フィールド調査でデータ収集の例
 - ベトナムで粗製肥料, マダガスカルで米流通
 - 歴史統計データを使った分析の例
 - 『帝国統計年鑑』の府県別データ
 - 官報の行旅死亡人公告データ (身長データの収集)
- 好きなエディタ: vscode > emacs
- 普段使っている言語: Stata, Python, R

はじめに

- 社会科学分野でも, 統計データ分析の重要性が増している
 - 日本には, 140年以上に渡る長期の政府統計がある
 - ただし, ごく近年まで紙媒体, よくて電子画像で提供.
 - 電子画像の統計表を, Excelファイルにテキストデータ化したい
 - 手入力? 膨大な時間と費用がかかる → 21世紀なので機械で
1. どこまで機械的に統計表を情報抽出できるか紹介
 2. その過程で出会った, 元祖「ネ申Excel」を観賞する

- はじめに
- OCRの活用と課題
- ocrstatsの概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

OCRの活用と課題

- OCR (optical character recognition: 光学的 文字認識)
- 統計表をOCRする際のポイント
 1. 文字(列)を正確に認識する
 2. 文字(列)を正確な位置(セル)に配置する

表レイアウト解析

		キロワット					
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

A B C D E F G H

1	2	3	4	5	6	7	8	9	10
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0		
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5		
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3		
115,250	265,434	7,409	2,617,500	6.5</td					

統計表をOCRにかけた例

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762
715,021	2,251,597	62,078	19,374,221	92.7	195.8	1083.4	9,324.3	7,828	8,881	21,543	28,521	151	5,556
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,069	6,743	5	534
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	16	4,419
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1	3,489	3,637	9,067	12,057	24	3,279
638,282	1,510,962	42,231	14,141,125	92.4	208.4	1426.7	9,548.4	6,457	6,860	21,916	29,894	11	380
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,340.8	726	787	3,006	3,937	106	163
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29
91,525	483,199	10,108	4,969,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20	—	—
2,844,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	73,964	107,273	402	20,925

43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1	
8,027	23,716	600	289,392	3.2	38.7	39.1	476.7	140	170	703	935	1	1	
6,993	21,307	344	222,590	2.6	26.9	23.9	247.5	49	51	161	214	3	190	
9,433	24,582	574	248,480	2.7	27.0	32.6	329.6	60	69	352	468	—	—	
20,854	47,811	1,203	488,888	5.0	51.4	70.7	814.6	323	330	1,067	1,419	—	—	
22,541	67,889	1,899	800,428	7.4	65.7	125.5	1,110.3	455	511	1,010	1,344	4	680	
26,509	71,049	2,089	670,634	5.6	53.2	84.0	792.9	505	540	1,444	2,029	7	901	
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2	164	169	425	566	—	—	
19,946	50,374	822	521,084	5.1	52.3	122.3	1,295.5	210	231	831	1,107	—	—	
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	309	319	1,324	1,819	—	—	
15,474	37,391	645	373,219	2.9	28.6	143.1	1,406.0	155	172	670	761	—	—	
24,509	65,290	1,177	606,807	4.8	44.4	202.0	1,860.5	207	209	359	411	—	—	
503,227	1,683,655	47,976	14,007,656	59.9	49.8	67,856	17,598	16,715	14,508	19,296	123	866	—	—
108,481	303,888	9,283	2,637,395	26.5	230.3	195.1	1,942.2	1,385	1,616	3,426	4,558	28	4,690	
39,754	101,454	2,705	1,144,349	5.4	60.9	128.0	1,387.7	592	629	2,098	2,801	2	224	
29,878	58,997	1,586	509,560	7.5	65.0	221.5	1,914.2	301	308	739	982	—	—	
24,272	61,422	2,015	575,697	7.9	73.8	226.3	2,124.3	396	428	1,323	1,762	3	310	
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1	587	612	909	1,198	—	—	
42,699	122,338	3,394	1,471,821	8.4	101.6	143.2	1,739.3	434	522	1,235	1,643	16	4,419	
36,345	71,381	1,706	626,267	6.7	58.8	108.3	932.7	552	567	854	1,135	—	—	
25,508	51,296	628	46,4781	7.6	69.2	198.5	1,794.8	248	250	612	814	—	—	
14,951	36,319	961	349,412	6.2	59.7	125.2	1,205.5	147	156	382	509	—	—	
89,331	186,103	4,146	1,787,704	12.7	122.3	369.4	3,545.3	1,157	1,207	3,630	4,825	2	100	
96,899	239,550	5,558	1,996,245	12.2	101.7	765.0	6,382.3	1,858	1,938	4,199	5,585	19	3,172	
40,288	81,397	2,050	690,038	8.3	64.0	220.9	1,872.3	327	336	856	1,138	3	7	
78,286	210,985	5,949	2,806,086	17.3	234.4	711.5	9,664.8	992	1,040	4,001	6,053	—	—	
142,693	314,804	8,486	2,808,627	15.8	187.1	582.6	4,966.3	1,169	1,244	4,153	5,528	5	1	
368,219	877,496	26,399	7,388,938	40.3	339.6	7,582.9	26,851.9	3,612	3,882	11,740	15,625	6	379	
23,022	50,178	874	492,785	8.6	84.9	249.1	2,446.6	275	275	720	958	—	—	
26,062	57,499	630	584,680	7.6	77.2	185.1	1,882.3	409	419	1,302	1,730	—	—	
10,919	24,403	557	237,929	5.4	52.1	108.9	1,058.7	98	107	937	316	—	—	
19,455	39,467	960	311,597	5.3	41.5	90.6	715.0	69	71	423	561	—	—	
18,765	47,893	766	450,984	3.9	36.5	113.8	1,071.3	35	42	312	281	—	—	
53,258	118,261	3,056	1,056,175	7.3	65.0	227.1	2,028.1	488	525	1,987	2,643	166	163	
16,581	49,238	862	415,308	4.1	39.1	110.9	1,064.9	36	42	147	196	—	—	
11,559	27,064	542	265,576	3.7	36.7	99.8	980.4	101	100	353	469	5	25	
23,952	51,590	1,608	559,660	6.9	80.6	454.5	5,283.3	129	142	712	948	—	—	
27,364	69,048	1,112	516,397	6.3	47.2	202.4	1,513.6	204	216	984	1,308	—	—	
13,118	31,362	690	313,020	4.5	45.4	69.0	688.4	154	165	253	337	1	4	
20,918	50,457	985	567,093	5.6	63.0	125.3	1,408.1	120	139	1,219	1,621	—	—	
83,826	198,885	4,754	2,039,823	11.0	112.6	625.5	6,418.7	414	458	2,040	2,813	—	—	
13,973	44,156	700	450,331	6.5	66.3	275.8	2,812.5	108	117	1,034	1,477	—	—	
35,636	86,159	1,312	845,948	8.0	78.1	366.4	3,597.5	39	39	109	145	—	—	
8,001	29,738	585	376,157	2.3	29.4	63.9	808.1	89	89	188	250	—	—	
7,624	20,738	602	219,758	3.5	37.5	42.6	450.9	84	101	582	771	—	—	
21,548	53,066	1,170	500,536	3.9	36.6	88.1	830.7	220	245	1,222	1,625	7	4,802	
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20	—	—	
2,844,735	6,148,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	73,964	107,273	402	20,925	

統計表をOCRにかけた例

JUST PDF4 (!!)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1													777,			
2	43,641	128,961	3,4671	1,438,5311	7,81	87,2	21,2	236,0	524	3,214	4,1					
3	9	256,554	6,708	4,251,59	44,9	60,2	49,5	503,5	4,737	6						
4	4	2,251,59	2,520,59						1,433	7,82E31	1,21,548	15 151	5			
5	3	7	11	22,7	195,8	1,083,4				5,069	4					
6	4	244,913	5,72,07811	6,51	64,2	9,324,9				2,701	3, 16 24					
7	5	265,434	6,9,07811	7,71	80,8	368,4	3,270	11	1,234	3,4801	2,701	3, 16 24				
8	6	1,530,064,231	14,141,1251	208,4	1,026,79,548,4	6,4571	2,701	3, 16 24	2,701	3, 16 24						
9	7	118,978	273,2621,2601	2,471,393	5,31	48,2	137,2	1,240,8	726	3,006	3,166					
10	8	5,9901	179,064	4,021	1,695,0531	5,5	52,1	151,7	1,435,7	688	2,302	3,6				
11	9	191,5251	483,199	10,108	4,999,445	6,3	64,9	180,9	1,871,8	1,188	6,394	8,7				
12	10	6,670	11,0991	204	82,5521	2,1	15,5	70,7	526,1	2	15	20				
13	14	2,344,735	6,127,42	160,865	56,726,676,611,116	107,2	248,0	2,287,25	27,376	79,96	10,402					
14	15	43,641	128,961	3,4671	1,438,5311	7,8	87,2	21,2	23	524	3,215	4,275	935			
15	16	8,097	40,001	600	3,9	33,7	30,1	6,0	140	51	214					
16	17	6,993	21,307	344	22,590	2,6	26,9	25,9	247,5	49						
17	18	9,433	24,582	574	245,489	2,7	27,0	32,6	329,6	60	69	468				
18	19	20,854	47,811	1,203	485,888	5,0	51,4	79,7	814,6	323	330	1,419				
19	20	22,541	67,888	1,8991	600,428	7,9	65,7	125,6	1,110,3	4350	351	1,344				
20	21	26,500	71,04	2,088	670,834	6,6	53,2	84,0	792,9	506	610	2,029				
21	22	15,731	40,987	951	544,706	3,2	42,2	106,4	1,412,21	1,641	1,691	566				
22	23	19,946	50,374	822	321,084	5,1	52,3	122,3	1,265,5	210	221	1,107				
23	24	27,653	69,412	1,242	685,354	7,1	69,4	170,4	1,678,0	3091	319	1,819				
24	25	15,474	37,991	6451	373,219	2,9	143,1	11	1551	172, 209	761					
25	26	24,509	68,2901	1,177	4,8	28,6	200,2	40	207	414						
26	27	50,327	1,683,65	47,976	14,007,656	59,9	44,4	10,322,2	110,872,71	5,398	339	10,296				
27	28	108,481	303,888	9,265	2,637,395	26,5	230,3	1,952,1	16,942,2	1,385	3,426	4,558				
28	29	39,754	101,454	2,705	1,144,249	5,4	60,9	123,0	1,3	592,301	2,098	2,801,982				
29	30	29,878	58,997	1,586	7,5	221,5	87,									
30	31	24,2721	61,422	2,015	509,960	7,9	65,0	226,9	2,124,3	396	739	1,762				
31	32	21,346	43,561	1,108	388,194	6,8	60,9	159,9	1,425,11	587	909	1,198				
32	33	42,699	122,236	3,894	1,471,821	8,4	101,6	143,2	1,729,9	434	1,235	1,643				
33	34	36,345	71,381	1,706	626,267	6,7	58,8	106,3	932,7	552	854	1,135				
34	35	25,508	51,296	623	464,781	7,6	69,2	198,5	98	248,147	612	814				
35	36	14,951	36,3191	961	349,412	6,2	59,7	125,3	4	382	509					
36	37	28,8931	186,109	4,146	1,787,704	12,7	369,4	3,5	1,157,1,888	327,3,630	4,825	5,585	1,138			
37	38	30	78,286	210,984	5,949	2,866,0861	17,3	234,4	711,9	9,664,8	992	4,001	6,053			
38	39	31	142,693	314,804	8,469	2,808,627	15,8	137,1	582,6	4,956,3	1,169	4,158	5,528			
39	40	41	26,309	7,388,938	40,3	339,6	7,582,9	63,								
40	41	32	368,219	877,496	874	492,765	8,6	84,9	249,1	85	3,612	275,409	720	15,625	958 1,730	
41	42	33	10,919	24,4031	557	237,3291	5,4	52,1	108,9	1,0	237	316				
42	43	34	19,455	39,467	960	311,597	5,3	41,5	90,6	58,	423	561				
43	44	35	53,258	118,361	3,056	1,056,1751	7,3	36,5	113,8	2,028,1	488	1,987	2,643			
44	45	36	16,581	43,281	862	4,1	39,1	110,9	64,	36101	353	196	469			
45	46	37	11,550	51,596	542	3,7	36,7	99,8	9							
46	47	38	23,952	69,0481	1,668	599,660	6,9	80,6	454,5	1	129,204	154	712	948,1,308	337	
47	48	39	20,918	198,8851	985	567,093	5,6	63,0	125,3		120	1391	458	1,621		
48	49	40	83,825	4,754	2,039,922	11,0	112,8			414		2,814				
49	50	41	13,973	44,1561	700	450,231	6,5	66,3	5	2,8	108	117,39	1,417			
50	51	51	8,001	29,7381	585	376,157	2,3	29,4	63,9	8,1	89	290				
51	52	52	7,624	20,738	602	219,758	3,5	37,5	42,6	8,1	84	771				
52	53	53	21,5481	53,066	1,170	500,3361	3,9	36,6	88,1	830,7	220	245	1,625			
53	54	54	16,5701	11,099	204	2,1	15,5	70,7	526,1	2	2	20				
54	55	55	42,344	73,516	148,4201	160,86556	76,676,	11,6	107,2	248,012,287	1,25,227,3761	79,964	10,402			

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
2	43,641	128,961	3,4671	1,438,5311	7,81	87,2	21,2	236,0	524	3,214	4,275	1	bj 1			
3	94357	256554	6708	252059	4,5	44,9	60,2	593,5	1433	1511	4737	6409	15	1762		
4	715021	2251597	62078	19374221	22,7	195,8	1083,4	9324,3	7828	8881	21543	28521	151	5556		
5	115250	265434	7409	2617500	6,5	64,2	162,4	1601,8	1876	1975	5069	6743	5	534		
6	104552	244915	5723	2562869	7,7	80,5	137,3	1436,9	1234	1339	2701	3592	16	4419		
7	241469	543375	12715	4823396	10,7	94,8	368,4	3270,1	3489	3637	9067	12057	24	3279		
8	638282	1510962	42231	14141125	22,4	208,4	1026,7	9548,4	6457	6860	21916	29894	11	380		
9	118978	273262	6201	2471393	5,3	48,2	137,2	1240,8	726	787	3006	3997	166	163		
10	75990	179064	4021	1,695,053;	5,5	52,1	151,7	1435,7	588	623	2302	3062	6	29		
11	191525	483199	10108	4999445	6,3	64,9	180,9	1871,8	1074	1188	6394	8703	7	4802		
12	5670	11099	204	82552	2,1	15,5	70,7	526,1	2	2	15	20				
13	2344735	6127420	160865	56726676	11,6	107,2	248,0	2287,9	25231	27376	79964	107273	402	20925		
14	43,641	128,961	3,4671	1,438,5311	7,8	87,2	21,2	236,0	524	573	3214	4275	1	1		
15	8027	23716	600	289362	3,2	38,7	39,1	476,7	140	170	703	935	1	1		
16	6993	21507	344	225290	2,6	26,9	23,9	247,5	49	51	161	214	3	180		
17	9433	24582	574	248,489	2,7	27	32,6	329,6	60	69	352	468				
18	20854	47811	1203	188888	5	51,4	79,7	814,6	323	330	1067	1419				
19	22541	67889	1899	600,428	7,4	65,7	125,5	1110,3	1,655	351	1010	1344	4	680		
20	26509	71049	2088	670834	5,6	53,2	84	79,2	506	510	1444	2029	7	901		
21	15731	40987	951	544706	3,2	42,2	106,4	1414,2	164	169	425	566				
22	22	19946	50374	822	521084	5,1	52,3	122,3	1265,5	210	221	831	1107			
23	23	27653	69412	1242	683354	7,1	69,4	170,4	1678	309	319	1324	1819			
24	24	15474	37991	645	373219	2,9	28,6	143,1	1406,0	155	172	670	761			
25	25	24509	65290	1177	606807	4,8	44,4	200,2	1860,5	207	209	359	414			
26	26	503227	1683655	47976	14007656	59,9	98,6	13322,2	110873	5398	6175	14508	19296	123	866	
27	27	108481	303888	9265	2637395	26,5	230,3	195,1	1694,2	1385	1616					
28	28	39754	101454	2705	1144249	5,4	60,9	123	1							

統計表をOCRにかけた例(ocrstats)

		「キロサット」		「キロワット」		「キロワット」		「キロワット」		「キロワット」		「キロワット」		「キロワット」	
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1	1	1
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762		
715,021	3,251,597	62,078	19,374,231	92.7	195.8	1,083.4	9,324.3	7,828	8,881	21,543	28,521	151	5,556		
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,069	6,743	5	534		
241,469	944,918	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	16	4,419		
127,715	4,821,396	10.7	94.8	368.4	3,270.1	3,489	3,637	9,067	12,037	24	3,279				
638,275	151,962	42,231	14,141,125	92.4	208.4	1,026.7	9,548.4	6,457	6,860	21,916	29,894	11	380		
18,978	275,262	6,201	2,471,393	5.3	48.2	137.2	1,340.8	726	787	3,006	3,937	106	163		
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29		
191,525	483,199	10,108	4,969,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802		
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20				
2,844,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925		

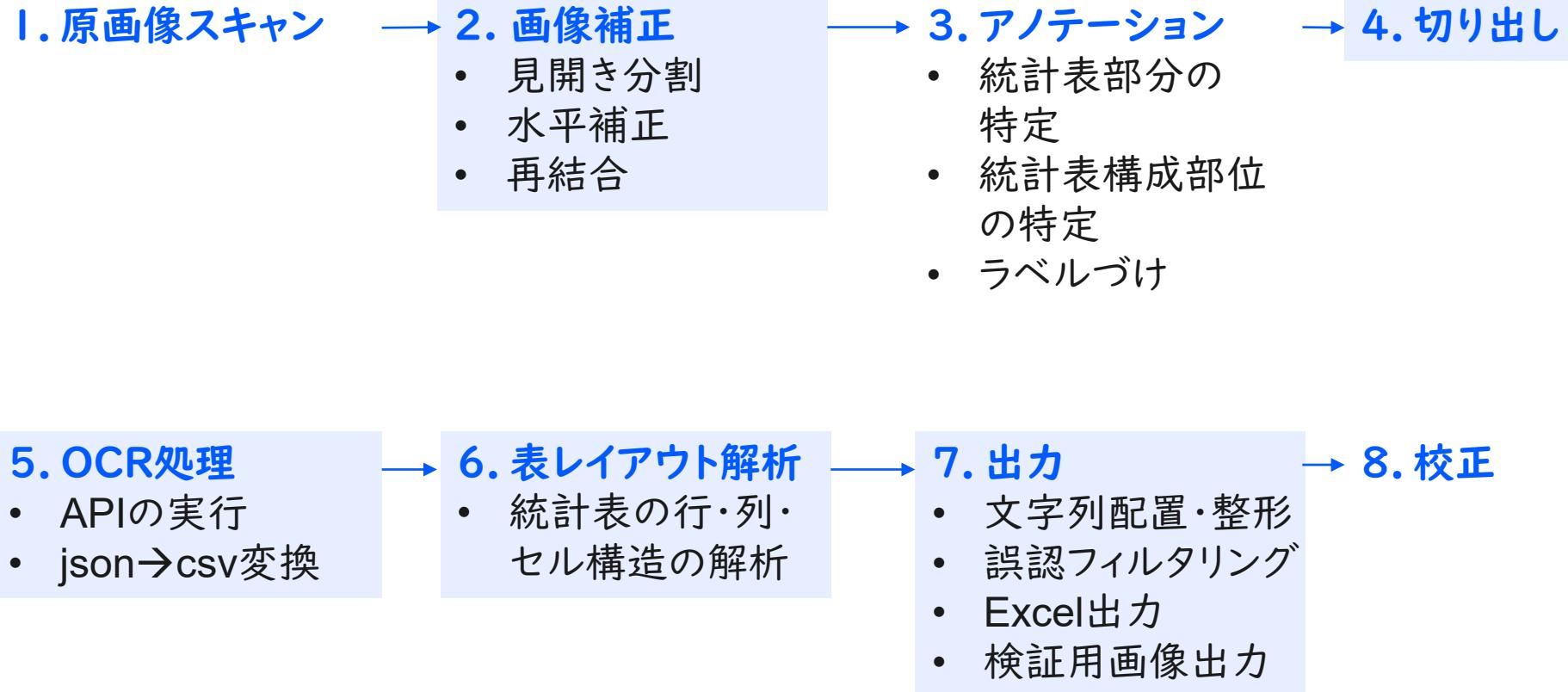
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P		
1	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1			
2	94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409	15	1,762			
3	715,021	2,251,597	6,251,821	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1		
4	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,069	6,743	5	534			
5	241,469	944,918	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592	12	379			
6	18,978	275,262	6,201	2,471,393	5.3	48.2	137.2	1,340.8	726	787	3,006	3,937	106	163			
7	75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062	6	29			
8	191,525	483,199	10,108	4,969,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703	7	4,802			
9	5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20					
10	2,314,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925			
11	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1			
12	8,027	23,716	6,606	2,601,362	3.2	38.7	39.1	476.7	146	178	763	935	1	1			
13	6,993	21,587	344	222,590	2.6	26.9	23.9	47.5	49	51	161	214	3	180			
14	9,433	24,582	574	248,489	2.7	27.8	32.6	329.6	86	89	252	465					
15	28,854	47,511	1,283	488,888	5.8	51.4	79.7	814.6	323	338	1,067	1,419					
16	22,541	67,889	1,899	806,428	7.4	65.7	125.5	1,110.3	355	351	1,810	1,344	4	680			
17	26,509	71,649	2,688	678,894	5.6	53.2	84.0	782.9	506	548	1,444	2,929	7	991			
18	15,731	49,987	951	544,766	3.2	42.0	186.4	1,414.2	154	166	425	566					
19	19,946	57,374	822	521,688	5.1	52.3	122.3	1,265.5	218	221	831	1,187					
20	27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	683,354	7.1	69.4	1,678.0	389	319	1,819		
21	24,489	73,219	1,897	806,428	7.4	65.7	125.5	1,110.3	355	351	1,810	1,344	4	680			
22	30,754	101,454	2,705	1,144,349	5.4	60.9	193.0	1,387.7	592	629	1,886.5	2,907	209	359	414		
23	39,754	101,454	2,705	1,144,349	5.4	60.9	193.0	1,387.7	592	629	1,886.5	2,907	209	359	414		
24	29,878	101,454	2,705	1,144,349	5.4	60.9	193.0	1,387.7	592	629	1,886.5	2,907	209	359	414		
25	24,509	65,290	1,177	606,897	4.8	44.4	200.2	1,860.5	1,860.5	1,860.5	1,860.5	207	209	359	414		
26	503,227	1,683,655	47,576	14,007,656	59.9	94.8	129,223	110,572	1,683,655	47,576	1,683,655	47,576	133,222	110,872	7	991	
27	108,481	303,888	9,265	1,438,531	7.8	87.2	21.2	236.0	524	578	3,214	4,275	1	1			
28	30,754	101,454	2,705	1,144,349	5.4	60.9	193.0	1,387.7	592	629	1,886.5	2,907	209	359	414		
29	24,489	73,219	1,897	806,428	7.4	65.7	125.5	1,110.3	355	351	1,810	1,344	4	680			
30	24,242	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0	683,354	7.1	69.4	1,678.0	389	319	1,819		
31	21,346	43,561	1,103	388,194	5.4	52.1	151.7	1,435.7	588	623	1,201.3	1,408.8	155	172	670	761	
32	42,699	122,236	3,394	1,471,821	8.4	81.6	143.2	1,723.9	434	452	1,235	1,643	16	4,419			
33	36,345	71,301	1,766	626,267	6.7	58.4	186.3	932.7	552	567	854	1,135					
34	25,508	51,206	623	464,781	7.6	69.2	106.5	1,798.4	248	256	612	814					
35	14,951	36,319	961	349,412	6.2	59.7	125.3	1,265.5	1,265.5	1,265.5	1,265.5	147	156	382	569		
36	35,331	188,199	1,416	1,860,060	17.3	234.4	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060	1,860,060
37	95,899	239,550	5,558	1,995,245	12.2	101.7	101.7	765,96,382.3	1,858	1,888	4,199	5,585	19	3,172			
38	40,268	81,397	2,650	698,835	8.3	64.0	220.9	1,872.3	327	336	856	1,136	3	7			
39	78,266	216,985	5,949	2,866,866	17.3	234.4	711.5	96,648.0	992	1,048	4,501	6,953					
40	142,693	314,864	8,469	2,888,627	15.8	137.1	682.6	4,956.3	1,150	1,244	4,153	5,628	5	1			
41	368,219	877,498	26,399	7,388,938	46.3	339.8	7,388,938	7,582,963,851.9	3,612	3,852	11,749	15,625	6	379			
42	23,922	50,118	1,874	492,183	8.6	84.9	249,12,446.6	275	275	729	998						
43	26,862	57,499	630	584,689	7.6	77.2	185.1	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3	1,882.3
44	18,919	24,463	557	237,329	5.4	52.1	108.9	1,658.7	98	107	237	316					
45	19,455	39,467	960	311,597	5.3	41.5	98.6	715.9	69	71	423	561					
46	18,765	47,803	766	458,084	3.0	36.5	113,81,671.3	35	42	212	281						
47	53,258	118,261	3,656	1,056,175	7.3	65.8	227.1	2,928.1	488	525	1,987	2,643	168	163			
48	16,581	43,238	882	413,398	4.1	39.1	116.9	1,064.9	36	42	147	198					
49	11,556	27,054	542	265,976	3.7	36.7	99.8	986.4	101	108	353	469	5	25			
50	23,952	51,589	1,658	599,666	6.9	65.6	454.5	5,283.3	129	142	712	948					
51	27,364	69,648	1,112	516,397	6.3	47.2	282.4	1,513.6	294	216	984	1,368					
52	13,118	31,362	699	313,820	4.5	45.4	69.8	688.4	154	165	253	337	1	4			
53	26,918	50,457	985	567,693	5.6	63.8	125.3	1,468.1	126	139	1,219	1,621					
54	83,825	198,385	4,754	2,639,922	11.0	112.8	62,586,418.7	414	458	2,048	2,814						
55	13,973	44,156	700	459,231	6.5	63.3	275.8	2,812.5	198	211	1,934	1,477					
56	35,636	86,159	1,312	845,948	8.0	78.1	366.4	3,597.5	38	39	109	145					
57	8,091	29,738	585	376,157	2.3	29.4	63.9	688.1	89	89	188	259					
58	7,624	29,738	602	219,758	3.5	37.5	42.6	459.9	84	101	582	771					
59	21,548	53,066	1,170	500,536	3.9	36.6	88.1	808.7	226	245	1,222	1,625	7	4,802			
60	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20						
61</td																	

- はじめに
- OCRの活用と課題
- **ocrstats**の概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

ocrstatsの概要

- 何?
 - 統計表画像のテキストデータ化を体系的に支援するPythonツール
 - 大量の統計表を体系的にテキストデータ化することを想定
- 特徴・要件
 - (1)一括処理, (2)定型工程の自動化, (3)外部OCRの活用
 - 実用的な精度で(とりあえず)動いて使えるものを作る
 - 使える外部ツールは躊躇なく, 節操なく, なんでも使う
 - アノテーションツール, OCRエンジン

作業の流れ



※青枠はocrstatsで自動処理

原画像

165. 原動力、電壓、發電力別事業數及發電力 (全 國) 大正二年 緯

	電氣供給及電氣製造事業			白家用電氣工作物			官設及 電氣工作 物		
	電氣供給	電氣製造	電氣機器及 電氣機器販賣	時	第一種	第二種	總	合計	
發電力(単位:キロワット)									
原動力	水力	212,847	180	28,525	212,552	3,650	31,794	400	278,396
	汽力	53,134	3,300	50,425	115,849	11,842	62,655	14,571	204,917
	天然力	6,905		434	7,359	1,875	10,684	270	20,293
	計	273,936	3,340	88,384	365,500	17,367	105,193	15,241	503,541
	電氣力	50,737	3,870	35,557	90,164	1,495	17,147	3,937	112,744
電壓	低壓	464	2,680	6,335	49,177	2,367	29,796	6,650	56,662
	高壓	62,759	800	41,549	95,118	—	53,050	4,881	153,649
	特別高壓	220,703	—	40,500	281,263	—	28,687	3,400	233,290
計		273,936	3,490	88,384	365,800	17,367	105,193	15,241	503,541

二種以上ノ原動力ヲ併用スルモノ其ノ主なガガ

第二種自家用電氣工作物ノ事業數中電氣事業法ノ準用スルモノ開業三アリ又電氣力ノ一部ヲ使用シタル容量不明ノ開業數五ヲ合計

166. 事業別發電所數、發電機台數及其容量（全國）

167. 電氣供給 (地方別) 大正二年

地 方	電 灯						電 力					
	供給戸数	電燈發電量	瓦	瓦	人口百人付	面積一里方付	供給戸数	電 灯	瓦	瓦	電力電氣力	其他電力装置
	電	瓦	瓦	電	瓦	電	電	瓦	瓦	換	換	瓦
統計上区分												
北 海 道	43,641	128,961	3,467	1,438,531	7.9	87.2	91.5	236.0	524	573	1,047	4,275
東 北 地	71,030	191,251	2,671	18,918,197,241	22.7	195.8	108,419	9,394.3	1,422	1,311	4,737	6,490
東 北 地	13,500	295,453	7,466	6,17,500	6.5	6.2	6.4	1,601.8	7,828	8,881	11,641	28,521
東 北 地	109,552	244,913	5,723	2,569,869	7.7	80.5	137.3	1,439.6	1,234	1,339	2,703	5,569
東 北 地	241,450	543,537	12,715	8,483,303	10.7	9.4	8,483,303	3,670	3,489	6,367	9,067	12,057
東 京 地	338,399	151,968	42,321	14,141,151	22.4	208.4	10,975	9,543.8	4,657	4,867	21,125	3,270
東 京 地	75,788	273,049	6,021	2,471,100	5.8	5.8	5.8	1,000.0	562	511	1,457	3,006
東 京 地	75,788	174,194	10,109	1,049,445	6.3	6.4	18,019	1,378.1	1,074	1,188	3,934	8,703
東 京 地	5,670	11,039	82,553	2.1	15.5	70.7	529.1	2	2	15	29	
總 計	2,344,735	6,127,490	100,955	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273
東 京 地												
東 京 地	49,641	128,061	3,467	1,438,581	7.8	87.3	91.2	236.0	524	572	3,214	4,275
東 京 地	8,027	23,716	0.0	289,362	3.2	38.7	39.1	74.7	149	170	703	935
東 京 地	21,507	52,992	590	2,922,590	5.5	25.5	25.5	1,000.0	52	51	101	214
東 京 地	574	1,471	1,471	2,922,590	5.5	25.5	25.5	1,000.0	60	60	324	508
東 京 地	29,884	47,811	1,203	488,886	5.0	51.4	79.7	81.4	323	323	1,067	1,419
東 京 地	29,541	57,889	1,839	600,428	7.4	6.5	125.5	1,110.3	455	351	1,010	1,344
東 京 地	24,509	71,049	2,588	701,514	5.6	53.2	84.0	79.2	506	540	1,444	2,029
東 京 地	15,731	40,987	951	544,700	3.2	42.2	106.4	1,414.2	164	161	445	506
東 京 地	19,346	50,374	822	521,104	5.1	52.3	132.3	1,265.5	210	221	831	1,077
東 京 地	27,633	69,414	1,242	585,354	7.1	6.4	174.0	1,375.0	309	319	1,324	1,819
東 京 地	15,474	37,991	645	273,219	4.8	26.0	141.1	1,204.0	155	175	375	414
東 京 地	24,509	65,290	1,177	409,867	4.8	44.9	44.9	1,000.0	222	222	354	414
東 京 地	502,247	162,655	47,671	14,664,000	26.5	26.5	26.5	2,922.5	539	539	16,715	44,505
東 京 地	108,453	303,988	9,384	9,637,000	26.5	26.5	26.5	2,922.5	1,828	1,828	3,455	4,555
東 京 地	39,251	10,241	2,703	1,141,200	5.4	60.0	100.0	1,367.7	593	624	2,801	3,801
東 京 地	20,879	56,997	509	509,969	7.5	65.0	221.3	1,142.4	301	306	739	982
東 京 地	27,627	64,222	2,015	575,097	7.9	7.8	226.9	2,124.3	396	423	1,323	1,762
東 京 地	21,346	43,561	1,103	388,194	6.8	60.0	159.5	1,425.1	587	612	909	1,198
東 京 地	45,600	122,236	3,304	1,471,821	8.4	10.1	142.3	1,732.8	454	522	1,235	1,643
東 京 地	30,345	71,381	62,667	62,667	6.7	58.8	106.8	1,093.7	557	567	854	1,135
東 京 地	25,508	52,341	623	484,781	7.6	7.6	198.5	1,739.4	248	250	612	814
東 京 地	14,451	36,319	42,949	412	6.2	59.7	155.5	1,205.5	147	156	382	509
東 京 地	89,231	150,190	14,146	1,787,704	12.7	12.7	360.4	3,549.3	1,157	1,207	3,630	4,825
東 京 地	90,839	229,550	5,558	1,006,245	12.9	101.7	765.9	3,638.3	1,858	1,858	1,436	1,585
東 京 地	40,388	131,957	2,050	630,005	8.3	64.6	220.9	1,872.8	327	328	856	1,128
東 京 地	78,250	210,988	4,593	2,866,096	17.3	234.4	711.8	9,484.8	905	1,000	4,001	6,025
東 京 地	141,636	314,804	8,400	8,400,000	17.3	234.4	711.8	9,484.8	1,167	1,244	3,443	5,548
東 京 地	268,219	877,400	268,219	7,888,000	40.3	300	3,382.5	28,815.0	3,612	3,888	11,743	15,635
東 京 地	20,575	47,178	549	47,178	47,178	47,178	47,178	2,449.1	273	273	730	958
東 京 地	53,538	118,961	3,056	1,056,173	7.5	7.7	153.1	1,882.3	409	419	1,202	1,730
東 京 地	10,919	24,403	537	237,229	5.4	52.1	108.3	1,058.7	98	107	237	316
東 京 地	65,545	94,997	969	311,507	5.3	45.5	41.6	705.0	69	71	423	561
東 京 地	18,763	47,893	540	456,984	3.9	36.5	113.8	1,071.3	35	49	212	281
東 京 地	53,538	118,961	3,056	1,056,173	7.5	7.7	153.1	1,882.3	409	419	1,202	1,730
東 京 地	16,881	43,238	652	415,308	4.1	31.0	110.3	1,049.3	30	42	147	196
東 京 地	11,555	27,065	592	565,376	3.7	36.5	50.4	980.4	101	100	353	469
東 京 地	23,952	51,590	1,058	599,600	6.9	80.5	45.5	1,528.0	123	123	712	948
東 京 地	27,364	69,018	1,112	515,397	6.3	27.4	20.0	1,058.3	204	215	1,308	1,708
東 京 地	11,311	35,925	693	310,020	15.5	45.4	59.0	684.8	154	165	337	411
東 京 地	20,918	50,457	395	567,003	3.6	63.1	195.2	1,418.4	120	130	1,215	1,621
東 京 地	8,073	20,446	4,015	2,400,531	6.6	66.3	275.2	8,218.5	414	458	3,940	5,211
東 京 地	67,473	144,556	2,671	2,671,000	17.6	27.6	366.3	3,507.5	39	39	109	145
東 京 地	25,636	88,159	1,312	845,848	8.0	27.6	366.3	3,507.5	39	39	188	250
東 京 地	8,001	20,738	585	376,157	2.3	29.4	63.9	908.1	80	80	255	300
東 京 地	74,624	70,738	608	219,758	3.5	37.5	42.0	450.9	84	101	562	771
東 京 地	9,545	53,096	1,170	500,536	3.9	36.5	88.1	807.0	224	245	1,222	1,628
東 京 地	5,670	11,039	304	82,553	3.1	15.5	70.7	529.1	2	15	20	
總 計	2,344,735	6,145,420	100,955	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273

右表ハ一帯ニシテ二ニテノノ所要電力供給スルモノハ其設備ノ區別シ各所附載シニ得タ

電力ニハ電氣手帳ニ付シテノノ所要電力供給スルモノハ其設備ノ區別シ各所附載シニ得タ

者本表ハ一事業ニシテニ以上ノ府縣ニ瓦リ供給スルモノハ其設備ヲ區別シテ各所屬府縣ニ掲タル事業ニシテニ供給スルモノハ之ヲ合計シテ

原画像（見開き分割→水平補正→再結合後）

316 13. 雷氣事業及瓦斯事業

165. 原動力、電壓、發電力別事業數及發電力 (全 國) 大正二年 累

三種以上ノ原動力ノ併用ニセキノ其ノ末ニ水方ニ計上キヌ

電氣導管及電氣供給系統ノ事務監査課道ノ事務ノ主ノ下、供給ノ主機器ノモニタ全

第二種自家用電氣工作物ノ事業數中電氣事業法ニ準用スルモノ開業三アリ又電氣力ノ一部ヲ使用シタル容量不明ノ開業數五合計

166. 事業別發電所數、發電機箇數及其容量（全國）

	原動力				電 機				發電力							
	水力	汽力	瓦斯力	計	低壓	高壓	發電量	計	五十瓩	百瓩	五百瓩	一千瓩	一千瓩	二千瓩	以上	計
發電所數事業別 (大正二十年末)																
明治四十一年末	74	62	3	137	19	49	137	—	31	75	15	16	137			
同 四十二年春	94	74	2	170	16	128	26	170	85	95	20	32	170			
同 四十二年夏	109	88	4	201	16	149	36	201	—	38	113	22	28	201		
同 四十三年春	131	107	17	255	22	189	51	255	—	50	134	32	39	255		
同 四十三年夏	171	113	47	314	18	239	77	314	—	82	156	43	53	314		
大正元年春	215	118	91	434	24	204	104	434	—	124	185	49	66	434		
同 二年春	247	130	117	491	24	365	105	491	—	171	193	57	73	491		
發電所數事業別 (大正二十年末)																
電氣機械	216	83	100	408	13	315	80	408	—	165	169	40	43	408		
電氣機械	1	7	1	9	7	2	—	—	—	—	4	1	1	9		
電氣機械	30	40	7	77	4	48	25	77	—	6	29	13	39	77		
計	247	130	117	491	24	365	105	491	—	171	193	57	73	491		
發電所數 (単位:キロワット)																
明治四十一年末	—	—	—	—	63	947	15	325	56	80	147	26	16	325		
同 四十二年春	—	—	—	—	59	299	26	372	43	90	131	31	27	372		
同 四十二年夏	—	—	—	—	55	395	30	416	44	95	200	45	32	416		
同 四十三年春	—	—	—	—	72	384	48	514	47	107	245	63	52	514		
同 四十三年夏	—	—	—	—	73	409	50	625	48	128	302	79	88	728		
大正元年春	—	—	—	—	79	366	83	728	57	156	203	106	106	728		
同 二年春	—	—	—	—	82	655	114	851	29	166	395	121	137	851		
發電所容量 (単位:キロワット)																
明治四十一年末	—	—	—	—	6,668	54,881	12,950	74,429	1,569	5,334	28,576	15,950	28,100	74,439		
同 四十二年春	—	—	—	—	7,188	68,270	29,020	104,473	1,197	5,933	36,195	14,970	41,200	104,473		
同 四十二年夏	—	—	—	—	7,323	85,946	32,923	124,792	1,774	6,314	39,866	27,968	49,390	124,792		
同 四十三年春	—	—	—	—	10,389	117,123	48,673	176,071	1,451	7,074	49,277	39,398	78,381	176,071		
同 四十三年夏	—	—	—	—	11,085	134,859	53,459	234,128	1,508	8,683	62,471	50,115	121,233	234,128		
大正元年春	—	—	—	—	11,947	207,406	147,393	366,746	1,640	10,431	69,294	67,650	217,731	366,746		
同 二年春	—	—	—	—	13,174	230,826	243,768	479,768	1,631	11,028	72,191	75,081	266,930	487,768		

本表ノ發電所数ハ使用認可ノモノニシテ、同一發電所内ニ二種以上ノ電壓アリトキハ其高キ方計上ス

本表ノ發電所數ハ使用認可ノモノ、ミヲ掲ク〇同一發電所内ニ二種以上ノ電壓アルトキハ其高キ方ヲ計上ス

13. 電氣事業及瓦斯事業 317

167. 電氣供給 (地方辦) 大正二年

者本表ハ一事業ニシテ二以上ノ府縣ニ亘り供給スルモノハ其設備ヲ區別シテ各所屬府縣ニ掲ク
電力ニハ電氣者ニ付於ノ公々会アリ

表体のアノテーションと切り出し(クロッピング)

表題 電気供給 (地方別) 大正二年

地 方	電 燈						電 力					
	供給戸数	電燈部数	電 燈	電 燈	電 燈	電 燈	供給戸数	電 燈	電 燈	電 燈	其他電力設備	
			電 燈	電 燈	電 燈	電 燈						
東北	43,641	128,961	3,467	1,438,531	7,8	87.2	21,2	236,0	524	573	3,214	4,275
北関東	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
関東	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
関東	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,860	21,543	28,521
東北	118,978	273,262	6,201	2,471,300	5.3	48.2	137.2	1,240.6	726	787	1,026.7	7,548.4
東北	75,950	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
東北	101,521	483,190	10,108	4,909,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
東北	5,670	11,090	204	82,552	2.1	15.5	70.7	526.1	2	15	15.5	20
東北	2,844,755,612,740	160,865,56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273	402	20,925
東北	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
東北	94,357	256,534	6,708	2,520,591	4.5	44.9	60.2	595.5	1,433	1,511	4,737	6,400
東北	715,021	2,261,437	22.7	195,8	1,083,4	9,234.5	62,078	19,374,221	22.7	195,8	1,083,4	9,324.5
東北	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	7,408	2,617,500	6.5	64.2
東北	104,432	244,434	5,788	2,260,591	7.7	176.5	137.3	1,495.9	1,234	1,373.3	1,401.8	1,373.3
東北	78,467	207,375	11.5	1,820,591	10.7	91.5	3,399	2,701	3,399	16	4,419	3,279
東北	638,282	1,510,962	42,231	1,411,125	22.4	208.2	1,026.7	7,548.4	6,457	5,86		

OCR處理(文字認識) | json → csv

View Go Run Terminal Help

1915_p315_corrected_表体1.json - TeikokuTokei - Visual Studio Code

1915_p315_corrected_表体1.json.csv

text	confidence	bbox1.x	bbox1.y	bbox2.x	bbox2.y	bbox3.x	bbox3.y	bbox4.x	bbox4.y
6,743	1	1878	193	1993	193	1993	235	1878	235
8,881	1	1545	151	1655	147	1657	190	1546	194
9,324.3	1	1201	154	1348	154	1348	193	1201	193
5	0.98	2097	193	2131	193	2131	235	2097	235
12.7	1	823	1698	909	1698	909	1737	823	1737
122.3	1	940	1705	1047	1705	1047	1742	940	1742
147	1	1434	1653	1507	1653	1507	1695	1434	1695
156	1	1588	1656	1664	1656	1664	1695	1588	1695
382	1	1765	1658	1841	1658	1841	1698	1765	1698
509	1	1917	1656	1995	1656	1995	1698	1917	1698
534	1	2233	196	2309	196	2309	238	2233	238
1,234	1	1395	235	1507	235	1507	277	1395	277
1,339	1	1542	237	1658	233	1659	277	1544	280
1,436.9	1	1207	238	1348	238	1348	277	1207	277
104,552	1	42	232	198	232	198	274	42	274
12,057	1	1862	282	1987	282	1987	321	1862	321
2,701	1	1726	238	1839	238	1839	279	1726	279
244,913	1	225	235	379	235	379	274	225	274
3,279	1	2196	282	2309	282	2309	324	2196	324
3,637	1	1546	279	1656	279	1656	321	1546	321
42,231 14,141,12	0.95	426	324	765	324	765	363	426	363
5,723	1	444	235	559	235	559	277	444	277
7.7	1	844	238	906	238	906	277	844	277
10.7	1	823	282	906	282	906	319	823	319
16	1	2076	235	2128	235	2128	277	2076	277
80.5	1	958	238	1047	238	1047	277	958	277

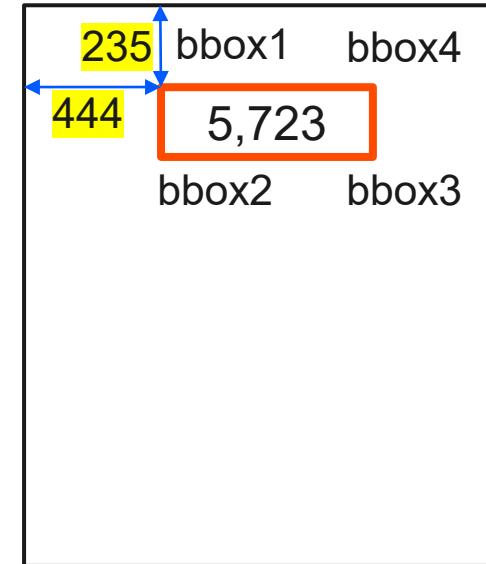
Ln 1635, Col 18 Spaces: 4 UTF-8 CRLF JSON

OCR処理

キロワット				キロワット				キロワット			
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0	524	573	3,214	4,275
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5	1,433	1,511	4,737	6,409
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3	7,828	8,881	21,543	28,521
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8	1,876	1,975	5,069	6,743
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9	1,234	1,339	2,701	3,592
241,469	543,375	12,715	4,823,396	10.7	94.8	273.2	3,270.1	3,489	3,637	9,067	12,057
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4	8,557	6,860	21,916	29,894
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8	726	787	3,006	3,997
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7	588	623	2,302	3,062
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8	1,074	1,188	6,394	8,703
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1	2	2	15	20
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9	25,231	27,376	79,964	107,273
										402	20,925

キロワット				キロワット				キロワット			
				3,467	1,438,531	7.8	87.2				
				6,708	2,520,591	4.5	44.9				
				62,078	19,374,221	22.7	195.8				
				7,409	2,617,500	6.5	64.2				
				5,723	2,562,869	7.7	80.5				
				12,715	4,823,396	10.7	94.8				
				42,231	14,141,125	22.4	208.4				
				6,201	2,471,393	5.3	48.2				
				4,021	1,695,053	5.5	52.1				
				10,108	4,999,445	6.3	64.9				
				204	82,552	2.1	15.5				
				160,865	56,726,676	11.6	107.2				

Text: 5,723
 Confidence: 1
 bbox1.x: 444
 bbox1.y: 235
 bbox2.x: 559
 bbox2.y: 235
 bbox3.x: 559
 bbox3.y: 277
 bbox4.x: 444
 bbox4.y: 277



表レイアウト解析(≒自動で罫線を引くこと)

「キロワット」									
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0		
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5		
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3		
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8		
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9		
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1		
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4		
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8		
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7		
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8		
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1		
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9		
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0		
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7		
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5		
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6		
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6		
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3		
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9		
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2		
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5		
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0		
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0		
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5		
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7		
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2		
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7		
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2		
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3		
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1		
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9		
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7		
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4		

「キロワット」									
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0		
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5		
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3		
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8		
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9		
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1		
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4		
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8		
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7		
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8		
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1		
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9		
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0		
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7		
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5		
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6		
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6		
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3		
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9		
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2		
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5		
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0		
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0		
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5		
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7		
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2		
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7		
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2		
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3		
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1		
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9		
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7		
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4		

表レイアウト解析1|OCR処理をする

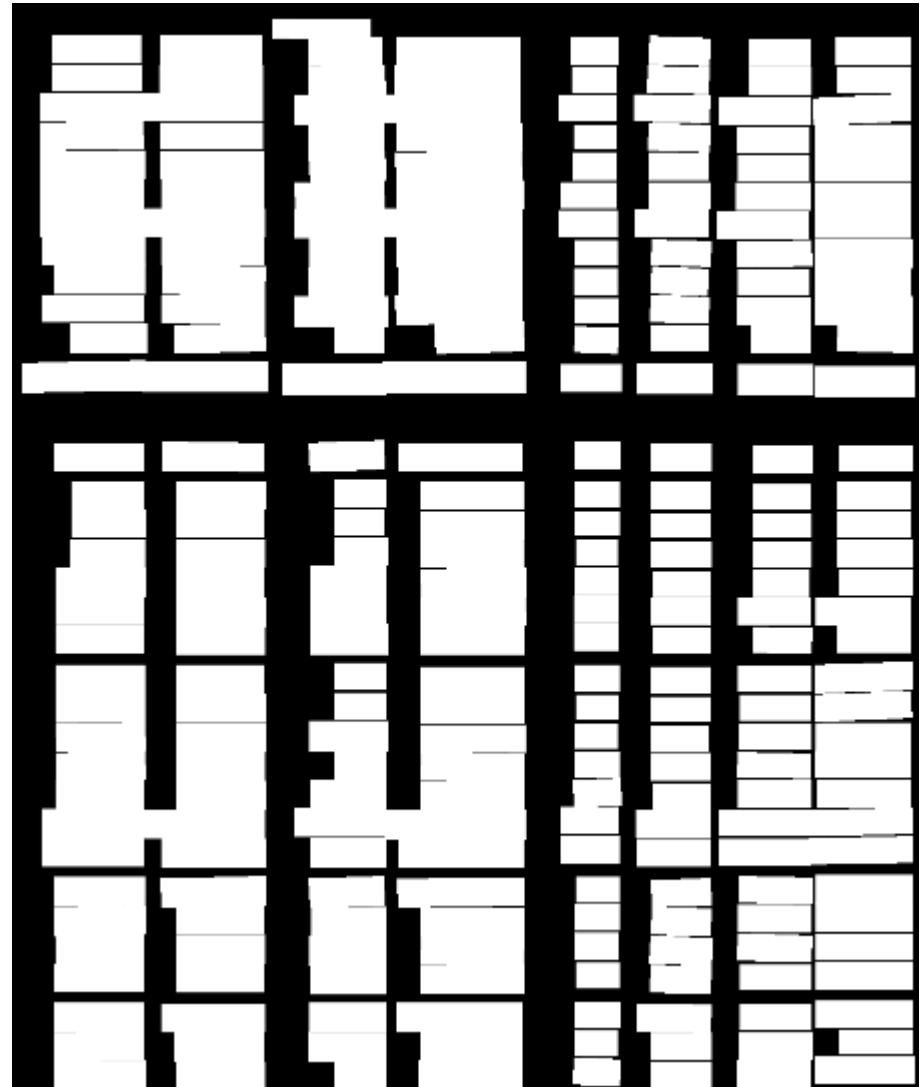
キロワット							
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

キロワット							
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4

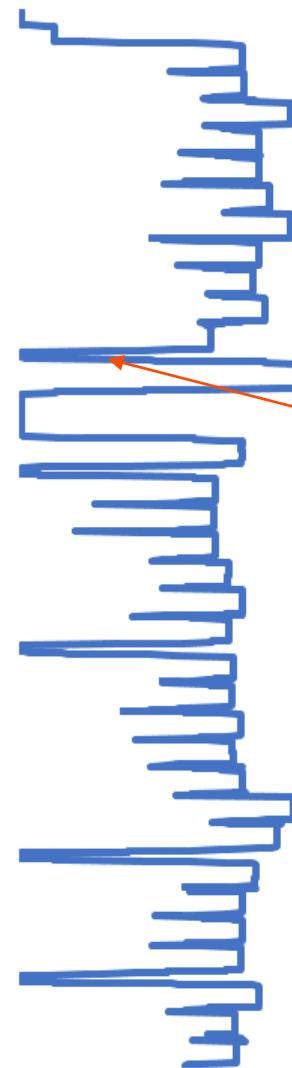
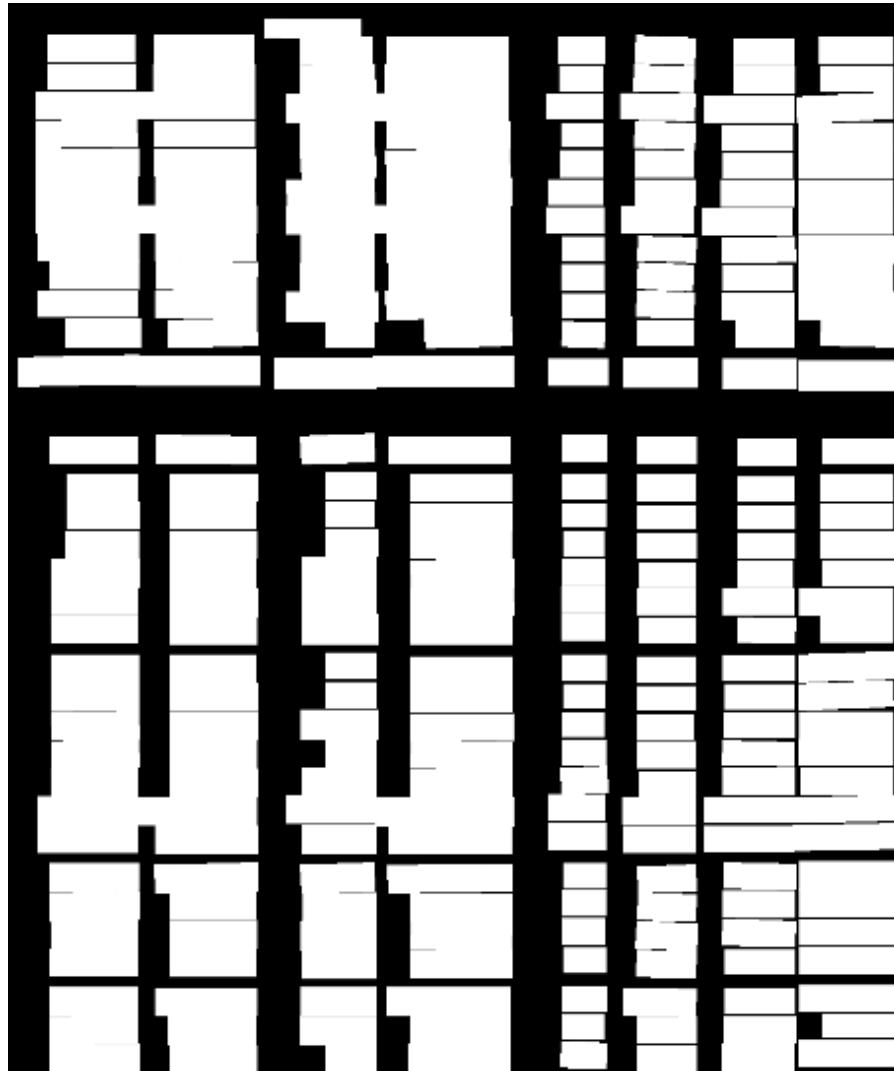
表レイアウト解析2|モノクロ画像を作成

キロワット							
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2	593.5
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4	9,324.3
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4	1,601.8
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3	1,436.9
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4	3,270.1
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7	9,548.4
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2	1,240.8
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7	1,435.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9	1,871.8
5,670	11,099	204	82,552	2.1	15.5	70.7	526.1
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0	2,287.9

キロワット							
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2	236.0
8,027	23,716	600	289,362	3.2	38.7	39.1	476.7
6,993	21,507	344	222,590	2.6	26.9	23.9	247.5
9,433	24,582	574	248,489	2.7	27.0	32.6	329.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7	814.6
22,541	67,889	1,899	600,428	7.4	65.7	125.5	1,110.3
26,509	71,049	2,088	670,834	5.6	53.2	84.0	792.9
15,731	40,987	951	544,706	3.2	42.2	106.4	1,414.2
19,946	50,374	822	521,084	5.1	52.3	122.3	1,265.5
27,653	69,412	1,242	683,354	7.1	69.4	170.4	1,678.0
15,474	37,991	645	373,219	2.9	28.6	143.1	1,406.0
24,509	65,290	1,177	606,807	4.8	44.4	200.2	1,860.5
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2	110,872.7
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1	16,942.2
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0	1,387.7
29,878	58,997	1,586	509,960	7.5	65.0	221.5	1,914.2
24,272	61,422	2,015	575,097	7.9	73.8	226.9	2,124.3
21,346	43,561	1,103	388,194	6.8	60.9	159.9	1,425.1
42,699	122,236	3,394	1,471,821	8.4	101.6	143.2	1,723.9
36,345	71,381	1,706	626,267	6.7	58.8	106.3	932.7
25,508	51,296	623	464,781	7.6	69.2	198.5	1,798.4



表レイアウト解析3|水平・垂直投影プロファイル

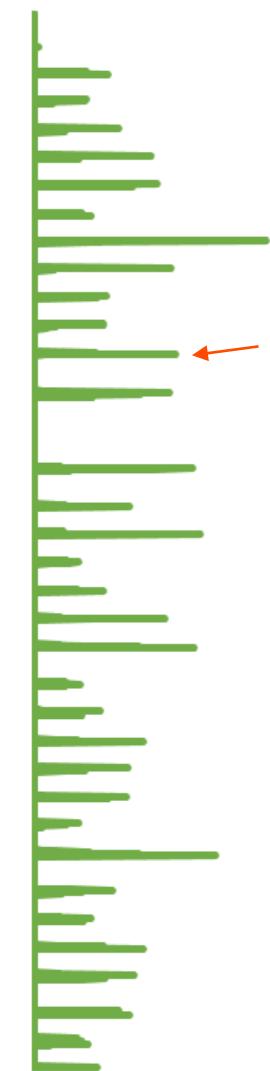
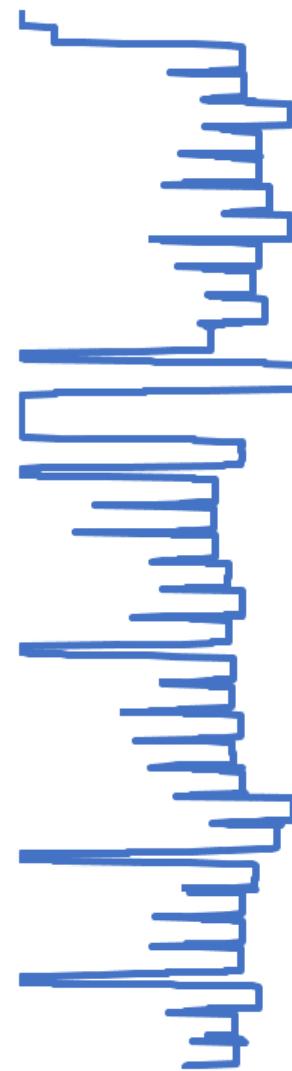
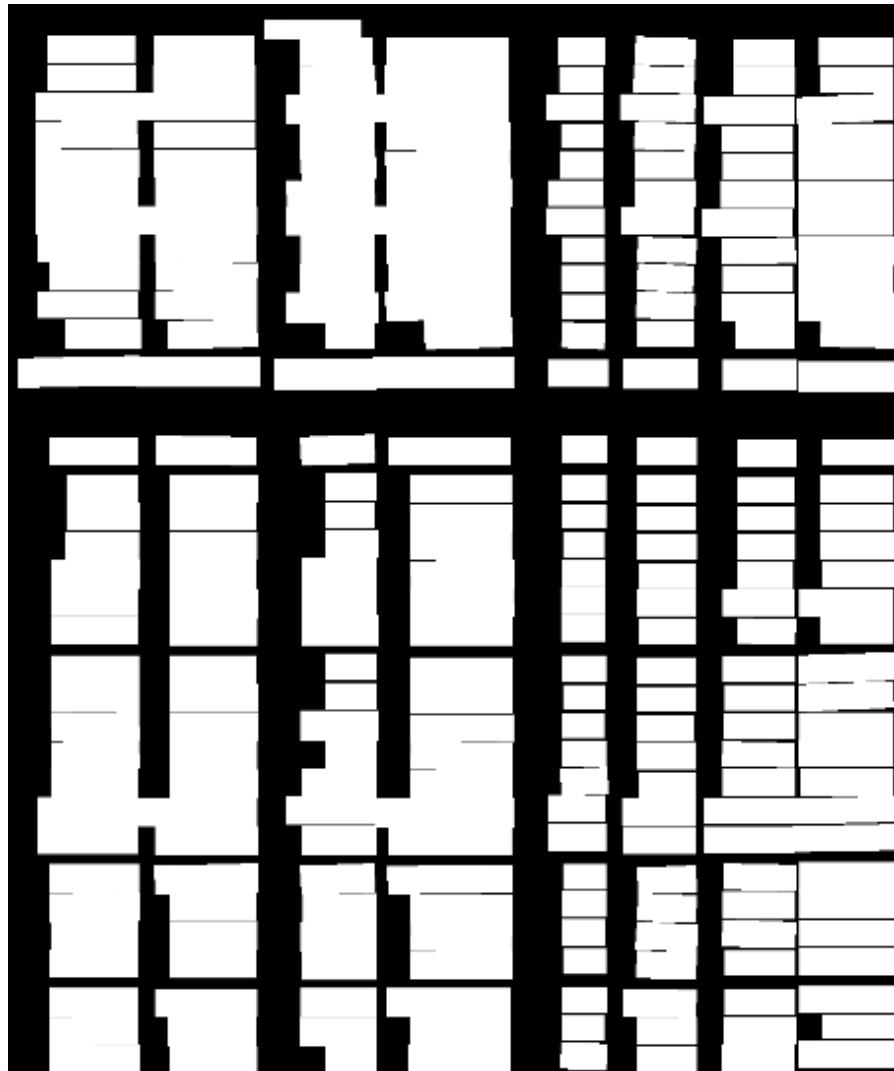


水平投影プロファイル
(projection profile)

行方向に, Y座標ごとに,
白いピクセル数を数えて,
ヒストグラムにする

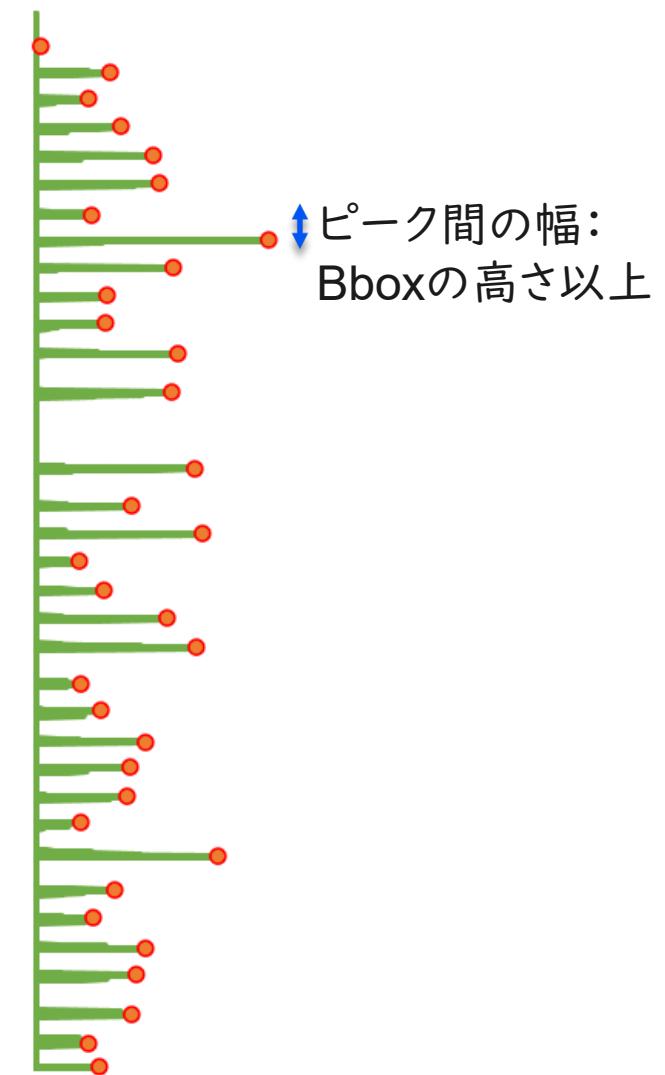
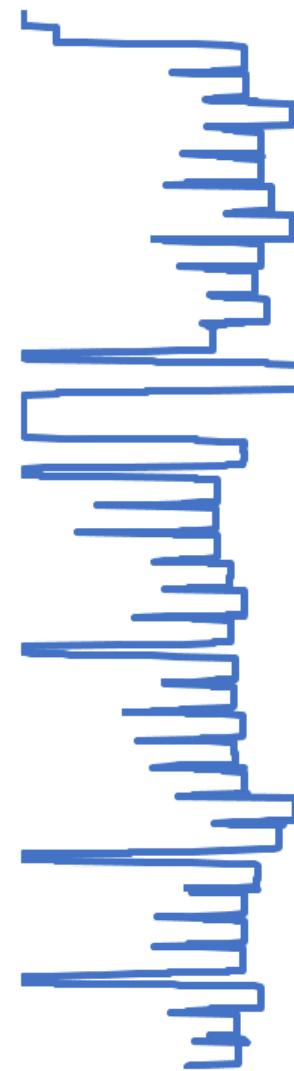
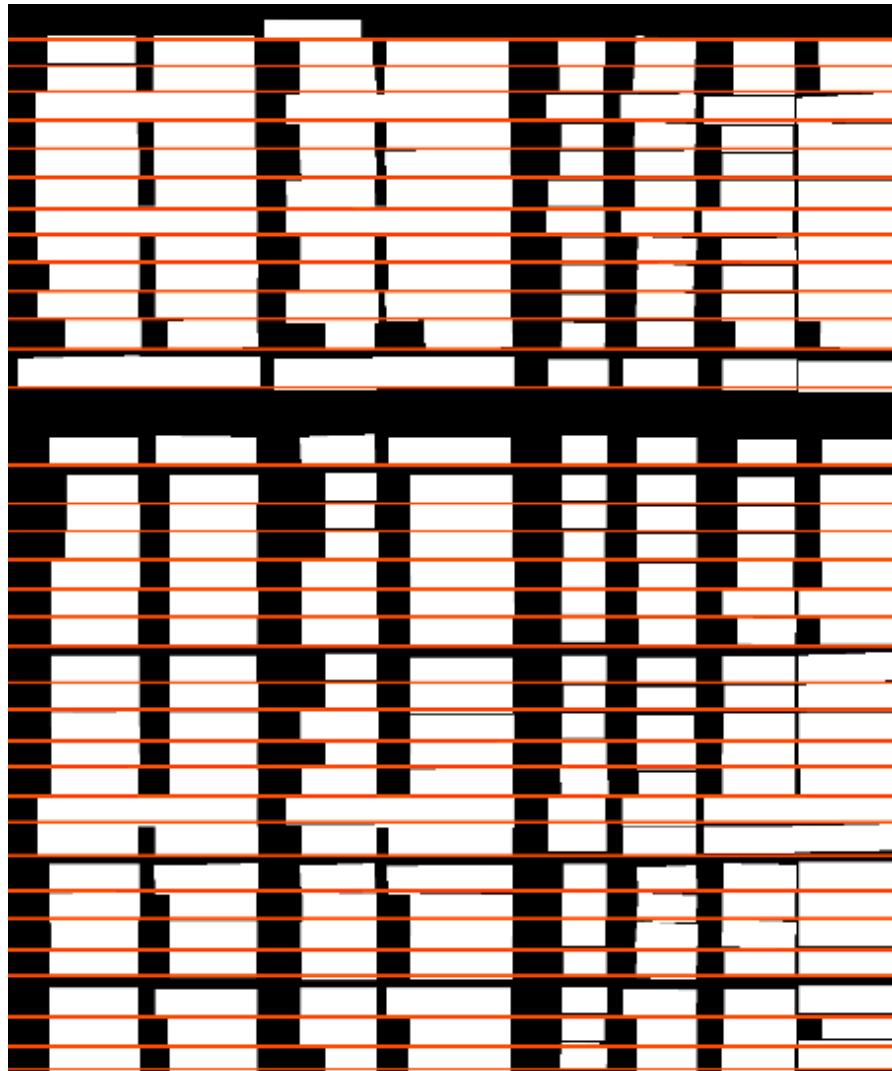
改行で急減

表レイアウト解析4|ヒストグラムの差分をとる

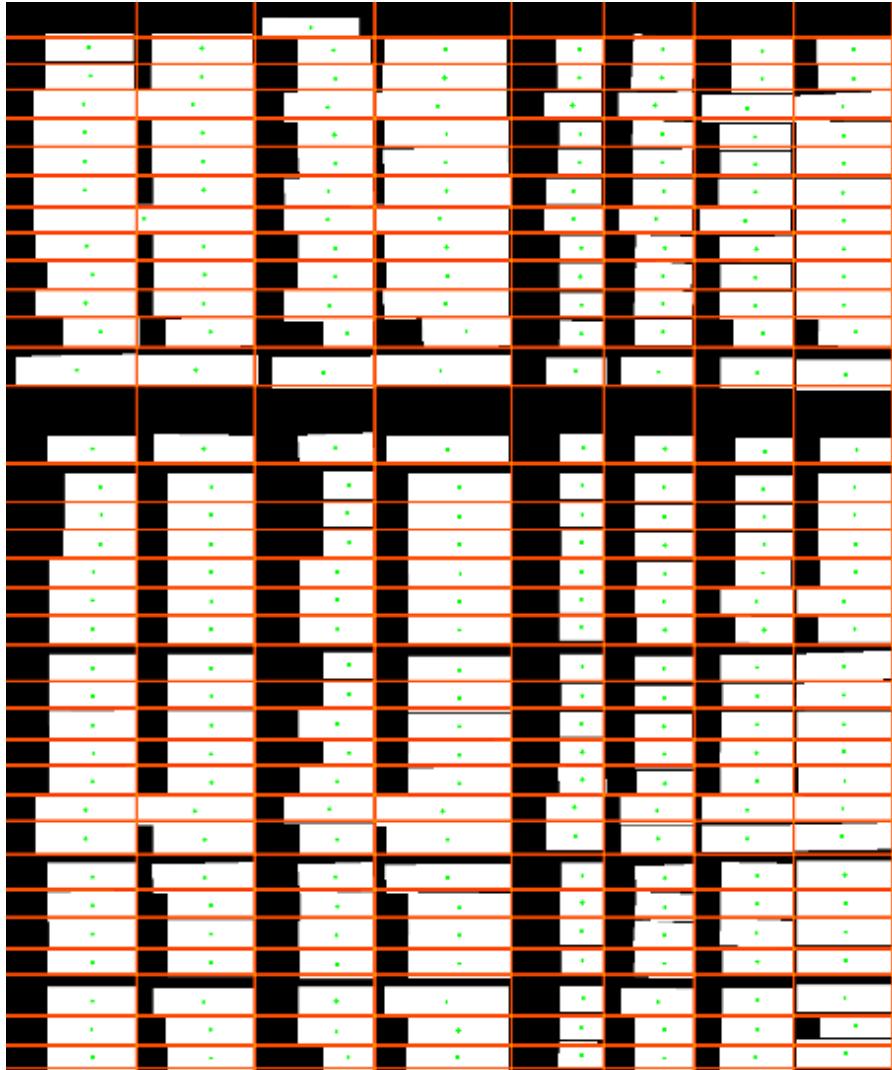


改行位置 =
減少幅が大きい
ところ

表レイアウト解析5|ピーク検出・改行位置の特定



表レイアウト解析6|文字列の配置



各boundingboxの
重心が含まれるセルに配置

統計表(Excel形式)と検証用画像の出力

キロット						
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
94,357	256,554	6,708	2,520,591	4.5	44.9	60.2
715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4
115,250	265,434	7,409	2,617,500	6.5	64.2	162.4
104,552	244,913	5,723	2,562,869	7.7	80.5	137.3
241,469	543,375	12,715	4,823,396	10.7	94.8	368.4
638,282	1,510,962	42,231	14,141,125	22.4	208.4	1,026.7
118,978	273,262	6,201	2,471,393	5.3	48.2	137.2
75,990	179,064	4,021	1,695,053	5.5	52.1	151.7
191,525	483,199	10,108	4,999,445	6.3	64.9	180.9
5,670	11,099	204	82,552	2.1	15.5	70.7
2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	248.0
43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
8,027	23,716	600	289,362	3.2	38.7	39.1
6,993	21,507	344	222,590	2.6	26.9	23.9
9,433	24,582	574	248,489	2.7	27.0	32.6
20,854	47,811	1,203	488,888	5.0	51.4	79.7
22,541	67,889	1,899	600,428	7.4	65.7	125.5
26,509	71,049	2,088	670,834	5.6	53.2	84.0
15,731	40,987	951	544,706	3.2	42.2	106.4
19,946	50,374	822	521,084	5.1	52.3	122.3
27,653	69,412	1,242	683,354	7.1	69.4	170.4
15,474	37,991	645	373,219	2.9	28.6	143.1
24,509	65,290	1,177	606,807	4.8	44.4	200.2
503,227	1,683,655	47,976	14,007,656	59.9	498.6	12,322.2
108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1
39,754	101,454	2,705	1,144,249	5.4	60.9	123.0
29,878	58,997	1,586	509,960	7.5	65.0	221.5
24,272	61,422	2,015	575,097	7.9	73.8	226.9
21,346	43,561	1,103	388,194	6.8	60.9	159.9

	A	B	C	D	E	F	G
1			「キロット」				
2	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
3	94,357	256,554	6,708	2,520,591	4.5	44.9	60.2
4	715,021	2,251,597	62,078	19,374,221	22.7	195.8	1,083.4
5	115,250	265,434	7,409	2,617,500	6.5	64.2	162.4
6	104,552	244,913	5,723	2,562,869	7.7	80.5	137.3
7	241,469	543,375	12,715	4,823,396	10.7	94.8	368.4
8			6,382,821,510,962		42,231	14,141,125	22.4
9	118,978	273,262	6,201	2,471,393	5.3	48.2	137.2
10	75,990	179,064	4,021	1,695,053	5.5	52.1	151.7
11	191,525	483,199	10,108	4,999,445	6.3	64.9	180.9
12	5,670	11,099	204	82,552	2.1	15.5	70.7
13	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	
14	43,641	128,961	3,467	1,438,531	7.8	87.2	21.2
15	8,027	23,716	600	289,362	3.2	38.7	39.1
16	6,993	21,507	344	222,590	2.6	26.9	23.9
17	9,433	24,582	574	248,489	2.7	27.0	32.6
18	20,854	47,811	1,203	488,888	5.0	51.4	79.7
19	22,541	67,889	1,899	600,428	7.4	65.7	125.5
20	26,509	71,049	2,088	670,834	5.6	53.2	84.0
21	15,731	40,987	951	544,706	3.2	42.2	106.4
22	19,946	50,374	822	521,084	5.1	52.3	122.3
23	27,653	69,412	1,242	683,354	7.1	69.4	170.4
24	15,474	37,991	645	373,219	2.9	28.6	143.1
25	24,509	65,290	1,177	606,807	4.8	44.4	200.2
26	503,227	1,683,655	47,976	14,007,656	59.9	498.6	
27	108,481	303,888	9,265	2,637,395	26.5	230.3	1,952.1
28	39,754	101,454	2,705	1,144,249	5.4	60.9	123.0
29	29,878	58,997	1,586	509,960	7.5	65.0	221.5
30	24,272	61,422	2,015	575,097	7.9	73.8	226.9
31	21,346	43,561	1,103	388,194	6.8	60.9	159.9

校正

地方	電 燈							
	供給戸數	電燈箇數	換算電氣力	燭光數	人口百二付		電燈箇數	燭光數
					電燈箇數	燭光數		
統計區畫			「キロワット」					
北海道	43,641	128,961	3,467	1,438,531	7.8	87.2		
東北區	94,357	256,554	6,708	2,520,591	4.5	44.9		
東北區	715,021	2,251,597	62,078	19,374,221	22.7	195.8		
東北區	115,250	265,434	7,409	2,617,500	6.5	64.2		
東北區	104,552	244,913	5,723	2,562,869	7.7	80.5		
東北區	241,469	543,375	12,715	4,823,396	10.7	94.8		
東北區	638,282	1,510,962	42,231	14,141,125	22.4	208.4		
東北區	118,978	273,262	6,201	2,471,393	5.3	48.2		
東北區	75,990	179,064	4,021	1,695,053	5.5	52.1		
東北區	191,525	483,199	10,108	4,999,445	6.3	64.9		
沖繩縣	5,670	11,099	204	82,552	2.1	15.5		
總計	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2		
道府縣								
北海道	43,641	128,961	3,467	1,438,531	7.8	87.2		
東北區	8,027	23,716	600	289,362	3.2	38.7		
	6,993	21,507	344	222,590	2.6	26.9		
	9,433	24,582	574	248,489	2.7	27.0		
	20,854	47,811	1,203	488,888	5.0	51.4		
	22,541	67,889	1,899	600,428	7.4	65.7		
	26,509	71,049	2,088	670,834	5.6	53.2		
	茨城縣	15,731	40,987	951	544,706	3.2	42.2	
	栃木縣	19,946	50,374	822	521,084	5.1	52.3	
關東區	27,653	69,412	1,242	683,354	7.1	69.4		
	群馬縣	15,474	37,991	645	373,219	2.9	28.6	
	埼玉縣	24,509	65,290	1,177	606,807	4.8	44.4	
	千葉縣	503,227	1,683,655	47,976	14,007,656	59.9	498.6	

A	B	C	D	E	F	G	H	I
統計区画	府県	電灯	電灯	電灯	電灯	電灯	電灯	電灯
	供給戸数	電灯箇数	換算電気力	燭光数	キロワット	人口百二付	電灯箇数	燭光数
5	北海道	43,641	128,961	3,467	1,438,531	7.8	87.2	
6	東北区	94,357	256,554	6,708	2,520,591	4.5	44.9	
7	関東区	715,021	2,251,597	62,078	19,374,221	22.7	195.8	
8	北陸区	115,250	265,434	7,409	2,617,500	6.5	64.2	
9	東山区	104,552	244,913	5,723	2,562,869	7.7	80.5	
10	東海区	241,469	543,375	12,715	4,823,396	10.7	94.8	
11	近畿区	638,282	1,510,962	42,231	14,141,125	22.4	208.4	
12	中国区	118,978	273,262	6,201	2,471,393	5.3	48.2	
13	四国区	75,990	179,064	4,021	1,695,053	5.5	52.1	
14	九州区	191,525	483,199	10,108	4,999,445	6.3	64.9	
15	沖縄区	5,670	11,099	204	82,552	2.1	15.5	
16	総計	2,344,735	6,127,420	160,865	56,726,676	11.6	107.2	
17								
18	北海道区	北海道	43,641	128,961	3,467	1,438,531	7.8	87.2
19	東北區	青森	8,027	23,716	600	289,362	3.2	38.7
20	東北區	岩手	6,993	21,507	344	222,590	2.6	26.9
21	東北區	秋田	9,433	24,582	574	248,489	2.7	27.0
22	東北區	山形	20,854	47,811	1,203	488,888	5.0	51.4
23	東北區	宮城	22,541	67,889	1,899	600,428	7.4	65.7
24	東北區	福島	26,509	71,049	2,088	670,834	5.6	53.2
25								
26	關東區	茨城	15,731	40,987	951	544,706	3.2	42.2
27	關東區	栃木	19,946	50,374	822	521,084	5.1	52.3
28	關東區	群馬	27,653	69,412	1,242	683,354	7.1	69.4
29	關東區	埼玉	15,474	37,991	645	373,219	2.9	28.6
30	關東區	千葉	24,509	65,290	1,177	606,807	4.8	44.4
31	關東區	東京	503,227	1,683,655	47,976	14,007,656	59.9	498.6

- はじめに
- OCRの活用と課題
- ocrstatsの概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

主なエラーのパターン

—		2		2
1		4		5
1		5		6
2		11		13
4		13		17
3		9		12
3		6		9
—		7		7

文字列範囲の認識不良

1	292	154,784	73,645	228,429
—	2,497	110,427	39,081	149,508
—	14	4,617	1,661	6,278
—	353	15,261	3,372	18,633
5	769	683,390	205,023	888,413
—	5,354	204,180	67,620	271,800
244	21,474	14,584,125	7,038,490	21,622,615
—	118,042	3,559,084	1,208,908	4,767,992
245	139,516	18,143,209	8,247,398	26,390,607
239	127,418	17,203,790	8,207,970	25,411,760
287	117,174	19,009,784	7,777,504	26,787,288
212	100,450	14,105,585	7,223,092	21,328,677
174	100,918	12,334,778	6,399,725	18,734,503
147	95,227	10,616,784	8,252,228	18,869,012

22	—
26	—
4	—
16	—
20	—
14	—
12	—
28	—
19	—
10	—
23	—
23	—
16	2
10	—
243	5

認識漏れ

表側は「縦読み」してしまう

總	計	縣
道	府	道
北	海	道
東北區	森	縣
關東區	手	縣
	田	縣
	形	縣
	城	縣
	島	縣
	木	縣
	馬	縣
	玉	縣
	葉	縣
	京	縣
	川	縣
	本	縣
	神	縣
	大	縣
	和	縣
		縣

總	計	縣
道	府	道
北	海	道
東北區	森	縣
關東區	手	縣
	田	縣
	形	縣
	城	縣
	島	縣
	木	縣
	馬	縣
	玉	縣
	葉	縣
	京	縣
	川	縣
	本	縣
	神	縣
	大	縣
	和	縣
		縣

そのままだと「縦読み」してしまう

	A	B	C	D	E
1		統計	區	畫	
2	北東關北東東近中四九沖	海北東陸山海畿國國州繩總道府	計		道區づ區區區區區區縣縣
3	北東北區	青岩秋山宮鯛	海		道縣縣縣縣縣縣縣
4	關東區	茨柄群埼千東神	奈	森手田形城島城木馬玉葉京川	縣縣縣縣縣縣府縣
5	北陸區	新富石w			縣縣縣縣縣縣縣縣
6	東山區	長岐滋			
7	東海區	山部愛三		鴻山川井 野阜賀 梨岡知重都庫阪民	縣縣縣縣
8	近畿區	京兵大奈和	歌	山	
9	中國區	鳥島岡廣山		取根山島口	府市三ス縣縣才オツウツカウツ一縣縣
10	四國區	德香愛高			縣m間縣
11	九州區	大福佐長熊宮鹿		島川媛知分岡賀崎本崎	縣縣入縣縣縣縣
12	沖		繩	兒島	縣

表頭の入れ子構造は難しい

地方	電 燈						電 力						
	供給戸數	電燈箇數	換 算 電 氣 力	燭 光 數	人口百ニ付		面積一方里ニ付		供給 戸數	電 動 機			其他電力装置
					電燈 箇數	燭光數	電燈 箇數	燭光數		裝置數	換 算 電 氣 力	換 算 馬 力	裝置數

地方	電 燈						電 力						
	供給戸數	電燈箇數	換 算 電 氣 力	燭 光 數	人口百ニ付		面積一方里ニ付		供給 戸數	電 動 機			其他電力装置
					電燈 箇數	燭光數	電燈 箇數	燭光數		裝置數	換 算 電 氣 力	換 算 馬 力	裝置數

	A	B	C	D	E	F	G	H	I	J	K
1	電	燈				電		力			
2	地方換		入口百ニ付		面積一方里ニ付	供給	電動	機			其他電力装置
3	供給戸敷電燈箇数電氣取引	燭光數電燈箇數	燭光數	電燈箇数	燭光數	戸數	装置数換電氣力	算か芳	装置数	換算電氣力	

ネ申Excelとは

- 「紙へ印刷することを前提に、セルの結合や罫線機能などをフルに使い、見栄えを優先して作ったExcelファイル」¹⁾
- 機械処理が難しく、データとしての再利用性が低い

「ネ申 Excel」問題

奥村晴彦@三重大学



<http://www.louvre.fr/en/oeuvre-notices/archaic-writing-tablets>

2)

区分	給与所得者数					給与額
	3月末	6月末	9月末	12月末	年間月平均	
千人	千人	千人	千人	千人	千人	百万円
平成18年度	54,752	55,013	54,065	53,396	54,506	200,034,648
19	54,957	55,311	54,417	53,768	54,817	201,272,200
20	54,672	55,269	55,000	54,739	55,124	201,317,695
21	56,232	54,909	54,035	53,884	54,967	192,474,249
22	55,817	54,887	54,367	54,153	54,792	194,372,236
23	55,169	54,688	54,459	54,273	54,647	195,799,665
	人	人	人	人	人	百万円
10人未満	10,010,414	10,022,094	9,900,419	9,627,198	9,890,052	30,235,033
10人以上	8,333,724	7,886,247	7,863,073	7,735,850	7,954,723	25,820,091
業	30人	8,725,936	8,668,139	8,630,664	8,653,029	8,669,395
所	100人	11,820,473	11,718,332	11,658,236	11,727,615	11,731,189
規	500人	4,039,942	4,061,021	4,072,913	4,081,849	4,063,946
模	1,000人	7,178,060	7,244,851	7,248,209	7,290,875	7,240,466
別	5,000人	5,060,628	5,086,941	5,085,593	5,156,124	5,097,320
	計	36,825,039	36,779,284	36,695,615	36,909,492	36,802,316
	合計	55,169,177	54,687,625	54,459,107	54,272,540	54,647,091
						195,799,665

- 1) <https://atmarkit.itmedia.co.jp/ait/articles/1612/26/news032.html>
- 2) <https://oku.edu.mie-u.ac.jp/~okumura/SSS2013slide.pdf>

総務省のガイドライン

- 総務省(2020)「統計表における機械判読可能なデータ作成に関する表記方法について」*

データ・項目の取扱い

- 1セル1データとなっているか (1-2)
- 数値データは数値属性とし、文字列を含まないこと (1-3)
- セルの結合をしていないか (1-4)
- スペースや改行等で体裁を整えていないか (1-5)
- 項目名等を省略していないか (1-6)
- オブジェクトを使用していないか (1-8)
- データの単位を記載しているか (1-9)

表の構成の取扱い

- データが分断されていないか (2-1)
- 1シートに複数の表が掲載されていないか (2-2)

「統計表における機械判読可能なデータ作成に関する表記方法」令和2年12月18日統計企画会議申合せ

ガイドラインのNG項目が、
約100年前の政府統計書で
ことごとく確認できる



「ネ申Excel」のルーツは
100年遡れる！

* https://www.soumu.go.jp/menu_news/s-news/01toukatsu01_02000186.html

数値・文字列・単位の混在、オブジェクトの利用

1セルに複数のデータが入力されている

	企業等数	売上金額（費用総額）
総計	1188389	391445 (355943)
A	50384	69565 (63883)
B	154138	50468 (45332)
C	86522	44098 (33210)
D	85983	22258 (20297)
E	580003	37652 (35016)
F	27456	15288 (14633)
G	119085	115499 (110735)
H	84818	36617 (32897)

氣 溫						
平均	累年平均 (観測年数)	最高 平均	最低 平均	最高 最低 ノ 較 差	最 高 月 日	最 低 月 日
度	度	度	度	度	度	度
24.4	24.3(10)	28.1	27.7	6.4	34.4(39.V.17)	10.1(34.II.12)
23.6	23.4(10)	28.2	20.1	8.1	37.9(36.VI.16)	8.2(34.II.13)
23.4	23.0(10)	28.6	19.5	9.1	36.2(34.VII.22)	3.0(34.II.13)
23.1	22.5(10)	25.8	21.3	4.5	33.1(41.X.6)	7.3(34.II.12)
22.6	22.0(10)	28.5	18.4	10.1	37.2(35.VIII.1)	-1.0(34.II.13)
22.1	21.6(10)	26.7	18.5	8.1	37.0(42.VIII.16)	-0.2(34.II.13)
21.9	21.5(10)	25.0	19.3	5.7	35.9(42.VI.9)	3.0(34.II.13)
23.4	23.2(10)	26.5	20.0	6.5	34.3(43.VII.12)	7.1(34.II.13)
22.4	22.0(20)	26.0	19.1	6.9	35.1(43.VII.21)	5.2(34.II.12)
21.2	20.8(10)	24.8	18.0	6.7	34.6(39.VIII.7)	3.1(34.II.12)
17.4	16.8(25)	21.4	14.0	7.4	36.2(26.VII.26)	-6.1(19.II.2)

「円」、「▲ (マイナス表記)」、「、(カンマ)」が文字列として入力されている

	単価	前回差分	生産台数
サンプル1	10,000円	130	12,000
サンプル2	9,100円	▲200	29,000
サンプル3	8,020円	▲350	37,000
サンプル4	7,500円	500	43,000
SUM関数	0	630	0
+ (加算演算)	#VALUE!	#VALUE!	121000

↑黄色の行は関数で合計を表示した例

年 次	數 量	價 額	數 量	價 額	數 量	ニ同々再掲セラル、姫アルモ貼ク度調査ニ據リ之ヲ掲載ス		
						真 鰯	春 黒 鰯	鰯
明治三十二年						48,538,428円	6,532,917円	9,688,513円
同 三十三年						48,672,087円	7,284	
同 三十四年						石 111		
同 三十五年						54,272,807円	7,019	
同 三十六年						石 50		
同 三十七年						62,313,363円	7,206	
同 三十八年						33,560,367円	5,715	
						39,748,566円	6,842	
						31,296,516円	5,692,690円	8,643,197円
						20,286,222	4,021,211	1,685,645
						20,286,222	4,021,211	11,185,188

オブジェクトを使用している

名 称	内 容	数 量
商品 A	サンプル 1 サンプル 2 サンプル 3	100 200 300

セル内改行

改行で体裁を整えている

薬剤名	出荷 本数	単価
鎮静剤 A-1	429	756
鎮静剤 A-2	321	648
鎮静剤 A-3	384	438
鎮静剤 A-4	408	775

横書き

縦書き

結果								人口萬=付	年判 次所 及番 裁號
闕席判決	拠棄認諾ニ 基ク判決	其他ノ終 局判決	取 下	和 解	訴狀差戻	其 他 ノ果	合 計	新受件數	
國									

入れ子構造

セルが結合されている

都道府県	エリア	市区町村	導入台数	増減数
	特別区部	千代田区	58406	11291
	特別区部	中央区	141183	18421
	特別区部	港区	243283	38152

§

東京都	特別区部	葛飾区	442913	327
	特別区部	江戸川区	681298	2331
	市町村	八王子市	577513	-2540
	市町村	立川市	176295	-3373
	市町村	武藏野市	144730	5996
	市町村	三鷹市	186936	853
	市町村	青梅市	137381	-1958
	市町村	府中市	260274	4788

§

市町村	青ヶ島村	-
市町村	小笠原村	-

48. 海外在留本邦人 (内地人)

關東州及南滿洲鐵道附屬地						支 那					
總 數		本業者		本業者十半從業者及家事使用人		總 數		本業者		本業者十半從業者及家事使用人	
男	女	男	女	男	女	男	女	男	女	男	女
1,027	821	665	81	362	740	443	307	344	106	99	201

紡織工業	製絲業
	縞絲紡織業
	絹絲紡織業
	麻絲紡織業
	毛絲紡織業
	撚絲業
	製錦業
	綿織物業
	絹織物及絹綿交織物業
	麻織物及其交織物業
機械及器具工業	毛織物及其交織物業
	其他ノ織物業
	染色及精練整理業
	莫大小業
	其他ノ紡織工業
	計
	名牌瓦斯及玻璃製造業
	氣體壓縮機及氣體製造業
	電氣機械器具製造業
	農業上木桿種精鍛造機用機器製造業
金屬工業	製造加工用機械器具製造業
	度量衡器及科學研究用機器製造業
	車輛製造業
	造船業
	其他ノ機械器具製造業
	計
	金屬精鍊業
	金屬材料品製造業
	鑄物製造業
	鑄物以外ノ金屬製品製造業
化學工業	鑄金業
	其他ノ金屬工業
	計
	陶磁器製造業
	硝子製造業
	セメント及石灰製造業
	煉瓦製造業
	其他ノ窯業
	製紙業
	人造紡絲製造業
工業	燐工製造業
	認製造業
	油脂類製造業
	セルロイド製造業
	醫藥及賣藥製造業
	工業製造業
	染料製造業
	塗料及顏料製造業
	石鹼化粧料品製造業
	人造肥料製造業
業	其他ノ化學工業
	計

均等割り付け

スペースで体裁を整えている

地域名	出荷本数	在庫本数
津□□市	429	756
四日市市	321	648
伊□勢□市	384	438
松□阪□市	408	775

職業	總數
1 農耕 = 從事スル者	
2 畜産 = 從事スル者	
3 蟻業 = 從事スル者	
4 林業 = 從事スル者	
5 漁業 = 従事スル者	
6 採炭 = 従事スル者	
7 採鑛 = 従事スル者	
8 石油鑛業 = 従事スル者	
9 土石採取 = 従事スル者	
10 烟業、土石加工 = 従事スル者	
11 金屬工業、機械器具製造、造船、運搬用具製造ニ從事スル者	
12 精巧工業 = 従事スル者	
13 化學製品、製造 = 従事スル者	
14 紡織工業 = 従事スル者	
15 被服、身贊品製造ニ從事スル者	

表側の省略

同じ名称を空白で省略している

薬剤名	出荷本数	在庫本数
鎮静剤 A - 1	429	756
2	321	648
3	384	438
4	408	775

		發送
		年度
大正	13	19,603,080
	14	20,954,554
昭和	1	22,584,405
	2	22,236,780
	3	24,564,036
	4	25,806,686
	5	26,380,272
	6	24,698,728
	7	24,703,934
	8	30,147,919

		普通	特殊	無料	合
			總數		
明治四十五年大正元年度		15,607	7,564	1,222	2
大正二年	年度	16,613	7,553	1,204	2
同三	年度	16,638	8,292	1,273	2
同四	年度	17,107	7,645	1,377	2
同五	年度	19,081	10,060	1,438	3
同六	年度	21,637	10,091	1,516	3
同七	年度	25,082	13,313	1,851	4
同八	年度	24,399	14,751	2,004	4
同九	年度	24,043	17,343	1,937	4
同十	年度	26,860	20,013	2,065	4
同十一	年度	27,878	21,357	2,121	5
同十二	年度	26,324	20,154	1,980	4

地	方
東京	明治三十七年度 同 三十六年度
神奈川	W W
埼玉	W W
千葉	W W
茨城	W W

長	崎	本部及支部
		管内區裁判所
佐	賀	本部
		管内區裁判所
福	岡	本部及支部
		管内區裁判所
大	分	
熊	本	
鹿	兒島	
宮	崎	
那	霸	本部
		管内區裁判所
崎	計	本部及支部
		管内區裁判所

レイアウト解析の難しさ

見開きで ページを またぐ

1ページに複数の表

工業及貿易 153

107. 肥料營業者及販賣價額(全國)

年次	年 末 营業者				製 造 販賣價額(円)					其 他	合 計
	製造者	輸入者	移入者	販賣者	動 物 質	植 物 質	鐵 物 質	調 合			
明治四十二年	18,516	448	414	35,618	6,626,327	6,494,810	6,864,274	9,618,632	18,072	59,617,115	
同四十三年	20,256	543	403	37,650	7,545,727	7,616,779	7,322,940	12,164,174	27,570	54,670,734	
同四十四年	21,794	606	409	39,719	7,580,854	7,979,777	8,000,988	22,291	29,291	40,677,353	
同四十五年	22,887	632	387	41,946	9,793,419	10,653,313	14,108,122	16,862,526	27,743	51,443,536	
大正二年	23,655	692	379	44,618	11,019,340	11,767,315	15,863,302	19,125,912	51,412	57,759,081	
同三年	23,700	678	381	45,681	10,562,157	11,977,941	16,314,055	17,481,472	54,168	56,389,771	
同四年	24,265	697	398	45,868	8,816,941	13,656,778	14,171,262	16,236,102	56,632	49,441,374	
同五年	24,335	687	413	45,470	9,489,888	18,865,735	19,841,201	12,995,735	85,703	61,841,914	
同六年	25,007	745	415	46,355	12,500,814	27,581,737	27,541,229	17,961,069	13,622	53,761,770	
同七年	25,775	834	423	47,865	16,965,197	49,208,400	35,218,305	17,771,13	130,096,713		
同八年	26,187	944	439	51,774	27,397,511	67,113,285	60,371,516	49,023,444	234,993	204,149,794	
同九年	25,048	1,052	476	52,432	25,359,428	52,855,075	58,793,075	54,564,051	1,332,611	175,900,955	
同十年	25,689	1,135	500	54,027	16,755,407	31,224,439	30,821,055	24,209,739	99,979	107,324,377	
同十一年	25,347	?	?	?	16,729,342	35,890,758	47,570,297	24,887,357	62,939	125,071,973	
大正二年	23,947	?	?	?	20,519,471	37,927,032	50,725,366	24,760,459	136,506	135,369,903	

大正二年以後「肥料取扱法施行規則」改正ニ依り製造價額ヲ換算

大正元年以後ハ開斜取締法施行規則ノ改正ニ依リ製造價

108. 內國製造煙草生產高 (種類大)

年 次	日有紙卷煙草				兩切紙卷煙草				葉 煙 草				劍 雄 草				價額合計		
	生 產	高	價	額	生 產	高	價	額	生 產	高	價	額	生 產	高	價	額			
大 正 二 年 度	5,378,503	21,331,783	7,201,162	7,134,162	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	57,049,292	54,135,292	85,039,504
大 正 三 年 度	5,942,356	22,224,200	7,145,088	7,144,437	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	59,207,297	54,101,703	81,112,695
大 正 四 年 度	5,963,006	22,735,594	5,187,346	7,520,193	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	57,561,251	47,042,732	50,096,175
大 正 五 年 度	6,359,200	20,132,170	2,342,435	8,791,194	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	46,425,246	49,736,799	84,712,993
大 正 六 年 度	5,309,513	24,648,245	2,638,269	11,507,269	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	6,498,473	55,396,593	102,167,638
同 七 年 度	11,256,318	52,408,476	31,10,211	14,627,940	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	4,495,193	47,482,765	139,678,290
同 八 年 度	14,460,047	78,099,993	2,932,244	14,745,565	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	4,068,792	77,044,489	175,122,294
同 九 年 度	18,775,071	105,540,363	3,088,753	21,09,199	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	12,172,626	28,746,949	102,894,237
同 十 年 度	20,307,375	107,063,947	2,882,271	3,075	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	7,468,403	83,841,230	292,923,357
同 十一年 度	21,305,543	128,899,445	22,305,127	825	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	130,106	139,828	7,007,341
同 十二 年 度	129,619,693	134,893,504	5,241,159	29,527,567	人 本	千 本	人 本	千 本	生 產	高	價	額	生 產	高	價	額	1,214	220,651	37,730,562
																6,214,825	75,772,952	224,714,962	

表側だけ
左ページ

178. 普通水解核苷酸脱水酶

108. 內國製造煙草生產高 (種類細別) 大正十三年

本表三於 τ 值低下之抑制劑

表の分断

第四卷 氧象 / 二 緒

西 楽 所	空 楽 ノ 錠 力		空 楽 ノ 錠 度		空 楽 ノ 錠 平							
	平 善	高 善	上月日	最 低	上月日	最 低						
八 本 丹	199.0	778.6	7 25	767.1	IX 25	12.7	82.0	VIII 25	8.7	9 25	18.1	80 6.9
石 井	708.0	116.5	1 35	246.7	VII 29	11.2	20.1	VII 29	-11.0	1 35	9.6	76 4.6
石 有	781.0	116.2	1 37	267.7	VII 27	10.7	29.3	VII 27	4.1	9 32	9.4	82 5.5
国 宮 本	709.6	775.7	1 27	741.0	XII 23	10.1	22.2	VIII 29	-12.2	1 27	8.8	80 6.4
九 本 有	705.5	774.8	1 26	747.1	XII 25	9.4	24.5	VII 21	-6.0	1 4	7.7	74 5.4
喜 道	708.2	227.1	1 27	745.1	XII 25	8.9	26.5	VII 21	-20.2	1 4	7.0	75 6.7
喜 道 郡	758.6	773.7	1 26	742.4	XI 6	8.6	31.0	VII 29	-23.0	1 4	8.1	84 7.7
札 佐	778.3	1 37	742.0	XI 6	7.2	26.6	VII 21	-12.7	1 4	7.2	78 6.8	
坂 本 川	705.0	775.6	1 26	744.2	XI 6	5.6	23.3	VIII 19	-13.6	1 29	6.4	82 7.5
十 本 有	708.3	775.6	1 27	749.0	XI 7	5.3	34.1	VIII 20	-31.8	1 27	6.6	78 5.8
喜 道	705.1	771.6	1 27	755.6	XII 21	7.6	—	—	—	5 29	6.2	74 5.2
京 宗	788.8	772.5	1 39	741.5	XII 6	6.3	28.6	VII 29	-21.6	1 16	8.5	87 8.7
喜 道	703.4	772.6	1 29	739.0	XI 7	6.5	26.5	VII 29	-21.0	1 26	6.6	73 6.6
喜 道	703.9	775.7	1 27	739.5	XI 7	6.2	21.6	VIII 20	-26.0	1 7	6.6	80 6.3
喜 道	705.6	772.5	1 28	739.5	XI 7	6.4	27.0	VII 21	-17.0	1 27	8.8	80 6.6
喜 道	700.1	771.5	X 15	728.1	XI 7	6.1	26.5	VIII 20	-14.9	II 8	6.2	82 8.0

第 五 表 氣 象 二

五 素 氣 象 二

主要漁獲物品目別價額 (單位千圓) (全國)

2,173	1,480	1,081	413	640	535	3,03	1,03	1,254	7,091	1,504
2,172	1,479	1,079	412	639	534	3,02	1,03	1,253	7,090	1,503
1,722	1,150	1,002	433	629	510	1,13	1,07	1,310	16,337	1,271
1,722	1,150	1,002	433	629	510	1,13	1,07	1,310	16,337	1,271
1,940	1,046	1,017	556	522	1,232	1,008	1,238	8,152	2,003	
8,201	981	1,205	537	510	984	1,177	1,474	8,935	1,484	
2,172	1,311	1,643	618	604	1,031	1,494	1,399	7,585	2,117	
3,000	2,994	2,941	1,000	999	2,999	2,999	2,999	1,160	3,751	
5,197	3,088	4,479	2,102	1,283	3,908	3,423	4,056	24,562	5,613	
6,113	3,506	3,818	1,767	1,295	1,691	3,875	4,326	22,918	4,419	
5,900	3,117	4,130	1,659	1,241	911	3,906	4,858	18,895	4,286	

網、鰐		貼	鮎	鰐	鮎	鰐	鮎	鰐	鮎	鰐	鮎	鰐	鮎
タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ	タラバ
512	6,632	3,888	3,655	1,881	1,335	1,561	3,609	5,923	16,933	4,183	1,064	1,064	1,064
5,123	3,989	2,659	2,559	1,579	1,401	5,815	5,250	17,745	5,949	17,745	5,949	5,949	5,949
512	6,632	3,888	3,655	1,881	1,335	1,561	3,609	5,923	16,933	4,183	1,064	1,064	1,064
594	8,667	793	1,317	920	323	210	643	2,878	1,073	1,073	1,073	1,073	1,073
1,208	8,667	899	1,344	795	207	707	5,451	1,123	1,123	1,123	1,123	1,123	1,123
925	1,035	913	1,035	407	203	637	4,559	1,059	1,059	1,059	1,059	1,059	1,059
1,555	1,241	887	1,274	1,069	448	163	654	5,168	1,034	1,034	1,034	1,034	1,034
1,697	1,094	938	1,338	979	435	169	650	4,481	976	976	976	976	976
1,188	987	1,107	1,459	858	540	144	533	3,243	1,069	1,069	1,069	1,069	1,069
1,170	1,210	1,229	1,378	1,049	340	144	533	3,243	1,069	1,069	1,069	1,069	1,069
2,133	1,524	1,523	2,195	1,410	807	200	971	5,817	1,388	1,388	1,388	1,388	1,388
2,420	2,917	2,031	2,978	2,074	1,259	346	1,262	6,191	1,830	1,830	1,830	1,830	1,830
3,532	3,173	3,004	4,071	3,433	1,657	657	2,050	7,530	2,968	2,968	2,968	2,968	2,968
2,479	2,704	2,704	2,704	2,110	2,110	709	709	1,397	3,398	3,398	3,398	3,398	3,398
2,709	3,421	3,933	2,929	2,755	2,338	605	1,908	10,302	4,167	4,167	4,167	4,167	4,167
1,969	1,363	1,887	2,438	2,355	975	409	2,372	12,867	4,155	4,155	4,155	4,155	4,155
2,329	2,695	1,599	4,019	2,810	630	467	2,364	25,651	4,043	4,043	4,043	4,043	4,043
底曳網(タラバ)		タラバ蟹	タラバ蟹	タラバ蟹	タラバ蟹	タラバ蟹	タラバ蟹						
1,682	—	562	1,009	870	—	573	139	16,088	68,971	68,971	68,971	68,971	68,971
1,798	—	601	501	1,189	—	688	199	13,032	75,296	75,296	75,296	75,296	75,296
2,225	—	663	494	1,392	—	791	214	21,650	82,019	82,019	82,019	82,019	82,019
2,019	—	717	1,353	1,349	—	739	245	17,565	85,731	85,731	85,731	85,731	85,731
2,181	—	738	1,373	1,373	—	732	245	17,565	85,731	85,731	85,731	85,731	85,731
2,291	—	638	974	1,603	—	838	208	21,350	95,063	95,063	95,063	95,063	95,063
2,293	176	849	585	1,888	2,255	937	248	18,811	94,836	94,836	94,836	94,836	94,836
2,493	216	1,026	565	4,429	2,337	1,240	321	20,213	127,235	127,235	127,235	127,235	127,235
3,411	293	1,069	521	4,257	2,337	1,115	303	29,213	127,235	127,235	127,235	127,235	127,235
5,169	604	2,019	515	4,423	5,887	1,349	578	35,304	35,304	35,304	35,304	35,304	35,304
6,485	611	2,287	1,038	3,594	3,004	1,540	737	50,910	240,834	240,834	240,834	240,834	240,834
7,899	1,045	3,591	547	2,724	4,903	1,111	677	51,518	270,204	270,204	270,204	270,204	270,204
7,909	1,045	3,591	547	2,724	4,903	1,111	677	52,734	255,236	255,236	255,236	255,236	255,236
7,696	1,434	3,584	613	4,884	4,647	1,437	703	52,734	255,236	255,236	255,236	255,236	255,236
8,244	1,538	3,584	613	4,879	4,645	1,399	923	49,512	251,079	251,079	251,079	251,079	251,079

粗密の妙

	梅 (石)	桃	梨	柿		全 國
				生 柿	干 柿	
大正二年	471,725	10,796,235	21,441,768	47,040,623	4,472,500	
同三年	436,009	10,463,266	20,847,354	38,911,224	3,707,462	
同四年	413,471	12,116,250	22,336,633	47,913,990	3,413,002	
同五年	341,744	12,964,142	24,635,620	40,039,291	2,215,552	
同六年	332,132	13,399,913	29,429,812	45,285,869	2,523,857	1
同七年	383,327	12,407,970	27,823,932	30,225,623	1,590,979	
同八年	411,927	12,809,160	27,726,480	43,617,962	2,251,680	
同九年	401,353	13,168,224	27,659,238	45,528,740	2,115,623	
同十年	405,823	11,065,483	30,128,501	49,615,279	2,857,540	
同十一年	383,794	11,321,854	30,230,244	47,609,824	2,865,061	1
同十二年	377,937	11,235,927	27,554,423	44,275,382	2,332,804	

地 方 別 (大正十一年)

地 方	金	銀	銅	鉛	錫	亞鉛	鐵	硫化銻	格魯謨	満鐵鑄	亞砒酸
1 北 海 道	915,600	205,989	452,410	—	—	—	2,097,237	—	12,324	40,058	—
東 北 地 區	2 青 岩 手 縣	6,845	14,777	90,639	—	—	—	—	—	4,347	—
	3 岩 手 縣	16,300	—	—	—	—	—	—	—	—	—
	4 宮 城 縣	218	20,738	—	76,720	—	299,159	—	—	—	—
	5 宮 城 縣	412,843	533,178	8,450,832	27	—	—	—	17,432	—	306
	6 山 形 縣	332	21,420	293,017	—	—	—	—	—	—	—
	7 福 島 縣	4,902	5,034	35,932	—	—	—	289,291	—	—	690
	8 茨 城 縣	2,131,135	842,947	6,009,178	—	—	—	—	531,052	—	—
關 東 地 區	9 木 馬 縣	178,321	711,435	9,415,391	—	—	—	—	—	2,048	290,214
	10 埼 玉 縣	—	—	—	—	—	—	—	—	1,450	—
	11 千 瓢 縣	—	—	—	—	—	—	—	—	351	—
	12 東 京 縣	—	—	—	—	—	—	—	—	—	—
	13 神 奈 川 縣	—	—	—	—	—	—	—	—	—	—
	14 福 井 縣	—	—	—	—	—	—	—	—	—	—
	15 新 潟 縣	611,315	218,439	14,705	—	—	—	—	—	—	—
北 陸 地 區	16 石 川 縣	—	—	1,446,557	—	—	—	100,666	—	—	—
	17 富 壱 縣	—	—	—	—	—	—	—	32,272	—	—
	18 福 井 縣	—	—	—	—	—	—	—	385	—	—
	19 山 梨 縣	14,800	—	—	—	—	—	—	131,981	—	—
	20 長 岐 縣	46,054	260,903	6,983	574,566	—	—	—	—	17,649	—
	21 阿 岐 縣	—	—	—	—	—	—	—	—	13,517	31,490
	22 静 知 重 縣	—	—	—	—	—	—	—	42,730	—	2,420
東 海 地 區	23 愛 知 縣	—	—	—	—	—	—	—	—	—	—
	24 三 重 縣	—	—	—	—	—	—	—	—	—	126
	25 滋 賀 縣	—	—	—	—	—	—	—	—	—	173
	26 京 大 県	—	—	—	—	—	—	—	—	34,955	—
	27 大 岐 縣	—	—	—	—	—	—	—	4,092	—	—
	28 兵 庫 縣	—	—	—	535,763	—	—	—	—	—	181,500
	29 奈 和 滋 縣	—	—	—	—	—	—	—	326,221	—	—
近畿 地 區	30 山 口 縣	—	—	—	—	—	—	—	—	—	—
	31 鳥 取 縣	—	—	—	—	—	—	71,268	—	122,851	—
	32 根 森 縣	7,200	77,527	831,131	—	—	141,141	—	1,208,803	7,564	—
	33 関 岩 縣	—	—	—	—	—	—	—	2,596	—	8,061
	34 廣 島 縣	—	—	—	3,000	—	564,900	—	697	—	343
	35 山 口 縣	—	—	—	—	—	—	—	—	—	42,614
	36 德 島 縣	—	—	—	—	—	—	—	—	—	—
四 國 地 區	37 香 川 縣	196,355	324,601	2,534,362	15,508	—	—	—	—	—	—
	38 愛 媛 縣	344,575	316,865	8,077,916	—	—	—	—	724,095	10,698	7,033
	39 高 知 縣	—	—	—	—	—	—	—	—	—	—
	40 福 岡 縣	2,924	60	—	21,892	—	4,047,710	—	—	—	473,195
	41 佐賀 縣	—	—	—	—	—	—	—	—	—	—
	42 長崎 縣	—	—	—	—	—	—	—	—	—	—
	43 熊 本 縣	—	—	—	—	—	—	—	—	—	—
九 州 地 區	44 大 分 縣	3,831,029	1,013,376	5,015,887	—	3,338	—	—	83	—	114,818
	45 宮 呉 縣	—	—	352,989	—	—	—	—	2,047	—	62,745
	46 虎 兒 島 縣	1,488,260	295,112	—	—	19,536	—	—	—	—	—
	47 沖 縄 縣	—	—	—	—	—	—	—	—	—	—
	總 計	10,209,068	4,862,401	43,623,979	691,713	558,637	5,052,910	7,861,105	2,991,929	153,437	157,700

あえて、ネ申Excelを擁護

- そこに美を見出す
 - 高度な活版印刷技術, ヒューマンな誤植, 罫線を多用した様式美
- データの利用可能性?
 - そんなことすんの、一部のマニアックな学者とかデータ屋だけ
 - 大多数は、集計された、見やすい「表」がほしい
 - ネ申Excelは「見やすい表」として、100年以上の伝統と合理性、美を兼備

認識範囲不良は短冊切り出しで改善

3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2	3.2	42.2	106.4	1,414.2
5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5	5.1	52.3	122.3	1,265.5
7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0	7.1	69.4	170.4	1,678.0
2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0	2.9	28.6	143.1	1,406.0
4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5	4.8	44.4	200.2	1,860.5
59.9	498.6	12,322.2	110,872.7	59.9	498.6	12,322.2	110,872.7	59.9	498.6	12,322.2	110,872.7	59.9	498.6	12,322.2	110,872.7	59.9	498.6	13,322.2	110,872.7
26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2	26.5	230.3	1,952.1	16,942.2
5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7	5.4	60.9	123.0	1,387.7
7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2	7.5	65.0	221.5	1,914.2
7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3	7.9	73.8	226.9	2,124.3
6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1	6.8	60.9	159.9	1,425.1
8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9	8.4	101.6	143.2	1,723.9
6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7	6.7	58.8	106.3	932.7
7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4	7.6	69.2	198.5	1,798.4
6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5	6.2	59.7	125.3	1,205.5
12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3	12.7	122.3	369.4	3,548.3
12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3	12.2	101.7	765.9	6,382.3
8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3	8.3	64.0	220.9	1,872.3
17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8	17.3	234.4	711.5	9,664.8
15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3	15.8	137.1	582.6	4,956.3
40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9	40.3	339.6	7,582.9	63,851.9
8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6	8.6	84.9	249.1	2,446.6
7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3	7.6	77.2	185.1	1,882.3
5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7	5.4	52.1	108.9	1,058.7
5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0	5.3	41.5	90.6	715.0
3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3	3.9	36.5	113.8	1,071.3
7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1	7.3	65.0	227.1	2,028.1
4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9	4.1	39.1	110.9	1,064.9
3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4	3.7	36.7	99.8	980.4
6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3	6.9	80.6	454.5	5,283.3
6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6	6.3	47.2	202.4	1,513.6
4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4	4.5	45.4	69.0	688.4
5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1	5.6	63.0	125.3	1,408.1
11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7	11.0	112.8	625.8	6,418.7
6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5	6.5	66.3	275.8	2,812.5
8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5	8.0	78.1	366.4	3,597.5
2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1	2.3	29.4	63.9	808.1
3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9	3.5	37.5	42.6	450.9
3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7	3.9	36.6	88.1	830.7
2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1	2.1	15.5	70.7	526.1
11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9	11.6	107.2	248.0	2,287.9

表側

		總	計
		道	縣
	北	海	道
東	青	森	縣
北	岩	手	縣
區	秋	田	縣
	山	形	縣
	宮	城	縣
	福	島	縣
關	茨	城	縣
東	柄	木	縣
區	群	馬	縣
	埼	玉	縣
	千	葉	縣
	東	京	府
	神	京	府
	奈	川	縣

水平罫線で画像を短冊切り出し
(表体画像の表レイアウト解析で
特定した改行座標情報を使う)

13	總計
14	道府縣北海道
15	青森縣
16	岩手縣
17	秋田縣
18	東北画山形縣
19	宮城縣
20	福島縣
21	茨城縣
22	栃木縣
23	關群馬縣
24	東埼玉縣
25	區千葉縣
26	東京府
27	神奈川縣
28	新潟縣
29	北富山縣
30	陸石川縣
31	固福井縣
32	東(長野縣
33	山岐阜縣
34	區滋賀縣
35	梨縣
36	東海江海塩岡縣
37	縣
38	三重縣
39	京都府
40	近兵庫縣
41	畿大阪府
42	區奈長縣
43	和歌山縣

- はじめに
- OCRの活用と課題
- ocrstatsの概要
- 元祖ネ申Excelとの邂逅
- 今後の課題と展望

現状・課題・展望

● 現状

- ocrstatsで、そこそこの精度の自動テキストデータ化を実現
- 現在、『日本帝国統計年鑑』のテキストデータ化進行中

● 課題

1. レイアウト解析の自動化

- 画像レイアウト解析（表の検出） + 表構成部位解析（表頭, 表側, 表体）

2. 表頭・表側の処理

3. OCRの改善: 活字, 旧字, 「統計表」データセットの機械学習

セマンティクスの利用

433 虞喜、宣公史、列傳及選

374. 現役陸海軍人及年僕

375. 屬有證道體

三 漢 舊 記

大正十二年六月廿九日、三管協三分署を新ト題目 341 ト定ム。昭和二年六月廿九日、三管署を新ト題目 341 ト定ム。

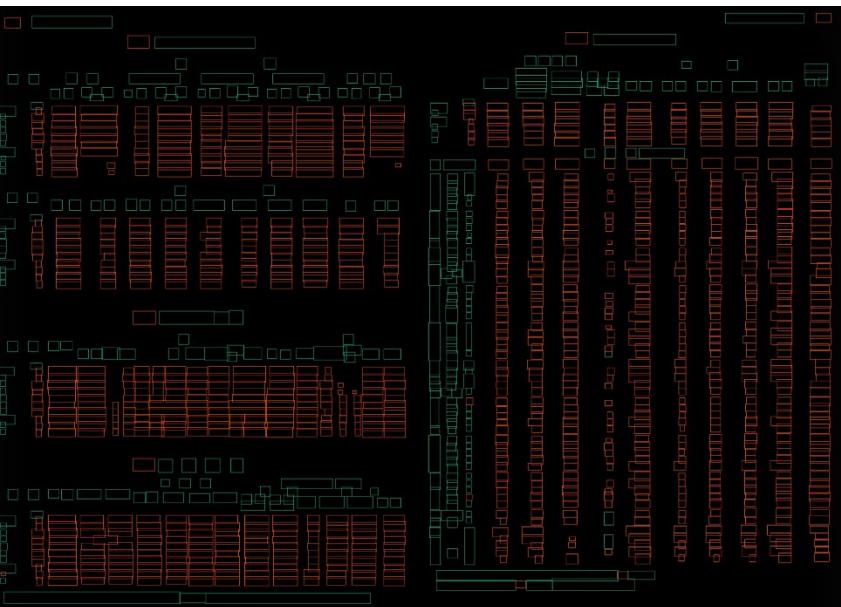
第100课 官公事，宣有及刑罚

理论与实践 | 第4期

國有鐵道職員

通 信 领 域

南洋監視局の監視係ニシタ越吉中ニハ監視係次1及ヒ監視 105 手合ニ。



現状・課題・展望

- 今後の展望
 - テキストデータ化する統計表の拡大
 - 『日本帝国統計年鑑』, 各省庁統計表, 府県統計書など
 - (政府) 統計レポジトリの作成
 - 統計の目録, メタデータ, 電子画像, テキストデータ, 作業情報の集約化
 - クラウドソーシング型プロジェクトの展開?
 - 校正をクラウドソーシング(例:「みんなで翻刻 <https://honkoku.org/>」)