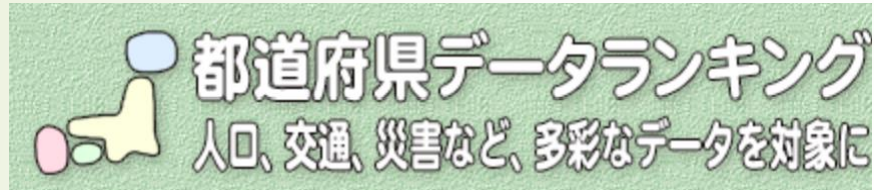




# オープンな統計データの利用

ネットで公開されている統計データを実験用データとして利用しようという場合に参考にしてください。

# 実験用データが入手できるサイトの例

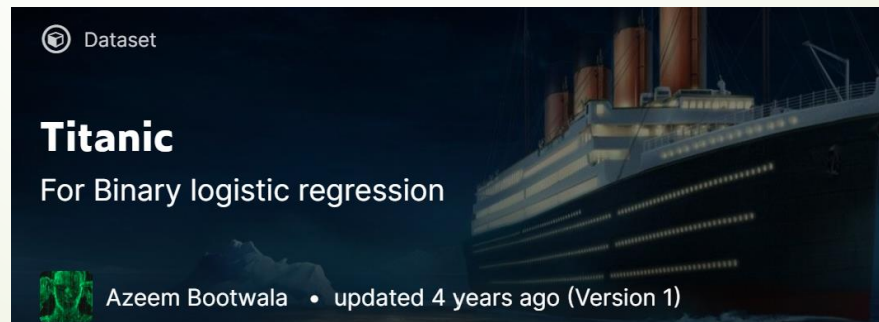


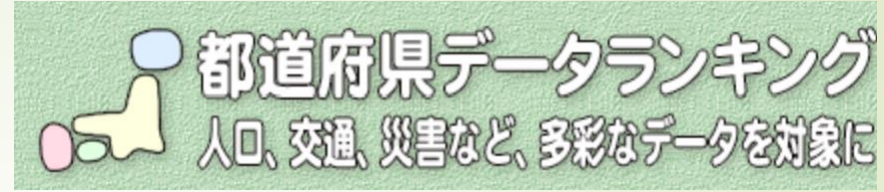
➡ [都道府県別データランキング](#)



➡ [e-Stat 政府統計](#)


➡ [Kaggle](#)





# データ加工の例

都道府県データランキングのサイトではブラウザ上に表データが表示されます。  
このデータを実験用のデータファイルにする手順を紹介します。



人口系   地理・気象系   事故・災害系   交通系   観光系   健康・生活系   社会 スポーツ・娯楽系   資源・エネルギー系   通信・IT系   政治・行政系   工業	
人口系	
<a href="#">人口・面積・人口密度</a> ...	人口、面積、人口密度（2010年から2020年まで各年10月1日現在）
<a href="#">人口変遷/大正から令和</a> ...	大正9年から令和2年まで5年間隔での人口変遷
<a href="#">自治体数(2000-2020)</a> ...	自治体数（市数、区数、町数、村数、総数）と前年増減数をランキング。
<a href="#">自治体数(1975-1999)</a> ...	平成の大合併以前（1975年～1999年）の自治体数をランキング。
<a href="#">市部・郡部</a> ...	市部(市と特別区)と郡部(町村)の人口、面積、人口密度、人口比率、面積比率

一番上の、「人口・面積・人口密度」を例に説明します。  
リンクをクリックして開いてみましょう。



## 都道府県別に並べ替え

2020年		10月1日現在			前年差	
#	都道府県	人口 (人) ↓	面積 (km2)	人口密度 (人/km2)	人口 (人)	面積 (km2)
1	東京都	14,064,696	2,194.03	6,410.44	121,840	-0.04
2	神奈川県	9,896,716	2,416.11	3,824.50	40,245	-0.19
3	大阪府	8,842,523	1,905.32	4,640.96	19,070	0.03
4	愛知県	7,546,192	5,173.07	1,458.75	-6,681	0.01
5	埼玉県	7,346,836	3,797.75	1,934.52	9,506	
6	千葉県	6,287,034	5,157.57	1,218.99	8,008	
7	兵庫県	5,469,184	8,401.02		5,575	

クリック

人口・面積・人口密度のデータの場合

最初は人口順に並んでいます。  
これを、道府県順に並べ替えます。



2020年		10月1日現在			前年差	
#	都道府県	人口 (人)	面積 (km2)	人口密度 (人/km2)	人口 (人)	面積 (km2)
1	北海道	5,228,885	78,421.39	66.68	-19,667	0.05
2	青森県	1,238,730	9,645.64	128.42	-7,408	
3	岩手県	1,211,206	15,275.01	79.29	-15,224	
4	宮城県	2,303,487	7,282.29	316.31	327	
5	秋田県	960,113	11,637.52	82.50	-5,855	
6	山形県	1,068,696	9,323.15	114.63	-8,361	
7	福島県	1,834,198	13,784.14	133.07	-13,752	0.24

# コピーしたいデータ範囲をマウスで選択



2020年		10月1日現在			前年差	
#	都道府県	人口 (人)	面積 (km2)	人口密度 (人/km2)	人口 (人)	面積 (km2)
1	北海道	5,228,885	78,421.39	66.68	-19,667	0.05
2	青森県	1,238,730	9,645.64	128.42	-7,408	
3	岩手県	1,211,206	15,275.01	79.29	-15,224	

42	長崎県	1,313,103	4,130.98	317.87	-12,102	-0.02
43	熊本県	1,739,211	7,409.46	234.73	-7,529	0.01
44	大分県	1,124,597	6,340.76	177.36	-9,834	
45	宮崎県	1,070,213	7,735.22	138.36	-1,864	-0.11
46	鹿児島県	1,589,206	9,187.06	172.98	-10,778	-0.02
47	沖縄県	1,468,410	2,282.59	643.31	14,226	1.59
全国		126,226,568	372,973.16	338.43	10,426	1.17



全国のデータは不要

- ◆エクセルのシートにコピペして不要な部分を削除。
- ◆見出しを1行に直す。
- ◆数値部分のフォーマットを通貨から数値に修正。

	A	B	C	D	E	F	G
1	2020年		10月1日現在			前年差	
2	#	都道府県	人口	面積	人口密度	人口	面積
3			(人)	(km2)	(人/km2)	(人)	(km2)
4	1	北海道	5,228,885	78,421	67	(19667)	0
5	2	青森県	1,238,730	9,646	128	(7408)	
6	3	岩手県	1,211,206	15,275	79	(15224)	
7	4						
8	5						

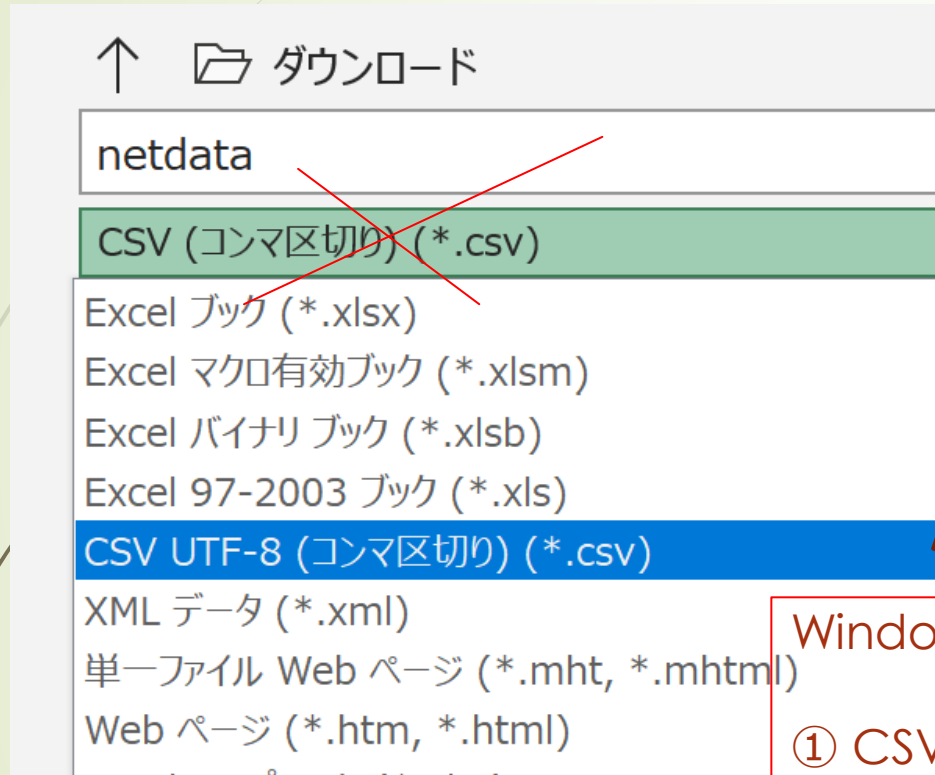
	A	B	C	D
1	都道府県	人口	面積	人口密度
2	北海道	5228885	78421	67
3	青森県	1238730	9646	128
4	岩手県	1211206	15275	79
5	宮城県	2303487	7282	316
6	秋田県	960113	11638	83
7	山形県	1068696	9323	115
8	福島県	1834198	13784	133



- (重要) 数値部分のフォーマットが通貨になっているので、数値に直しておく。
- 3桁おきにカンマ (,) が入っていると、プログラムでは文字として扱われてしまい、計算に使えない。



# CSV UTF-8 形式で保存



こちらを選ぶ

Windows の Excel では

- ① CSV
- ② CSV UTF-8

の2通りが選べるが、①はcp932, Shift-JISと呼ばれる形式で、ガラパゴスです。古い Office で②がない場合は、①を選び、プログラムを cp932用に修正しないといけません。