



データサイエンス入門B

第7回 e-Statの使い方

高田 美樹

目次

- ▶ データを集める・読む・説明する
- ▶ データの分類
- ▶ 定型データ
- ▶ データの可視化
- ▶ データの種類
- ▶ オープンデータ
- ▶ e-Statからのデータの取得
- ▶ データの可視化
- ▶ 指数化
- ▶ 成長率

データリテラシー

データを集める



データを読む



データを説明する



データの分類

データ

定型データ

非定型データ

時系列データ

クロスセクションデータ

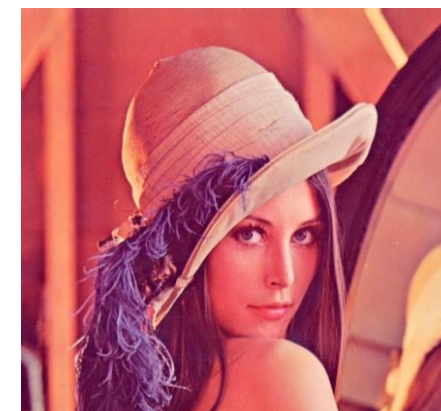
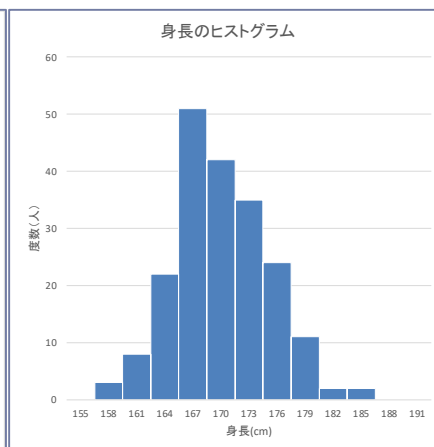
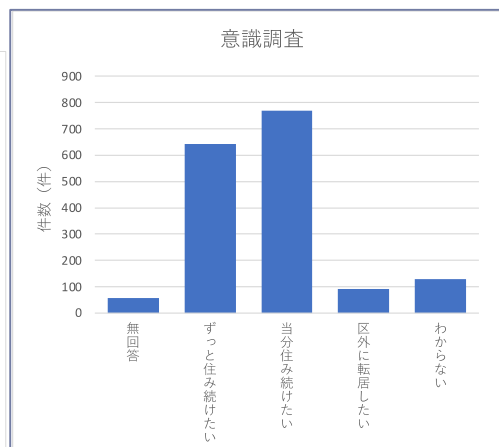
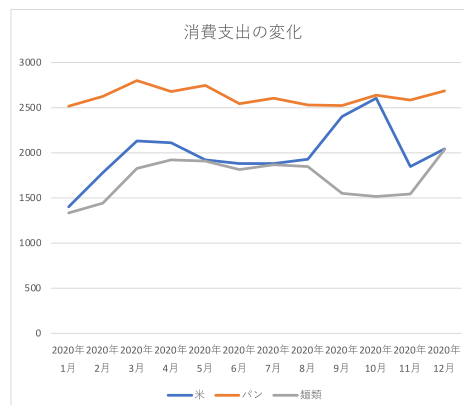
パネルデータ

画像・音・文書

質的データ

量的データ

時系列データ＋
クロスセクションデータ



定型データ

時系列データ

	A	B	C	D
1		米の生産量（日本）		
2	Year	Year	Unit	Value
3	1961	1961年	t	16160440
4	1962	1962年	t	16927100
5	1963	1963年	t	16648830
6	1964	1964年	t	16356100
7	1965	1965年	t	16126060
8	1966	1966年	t	16564000
9	1967	1967年	t	18782000
10	1968	1968年	t	18779010
11	1969	1969年	t	18200000
12	1970	1970年	t	16493000
13	1971	1971年	t	14148000
14	1972	1972年	t	15450500
15	1973	1973年	t	15778000
16	1974	1974年	t	15064000

クロスセクションデータ

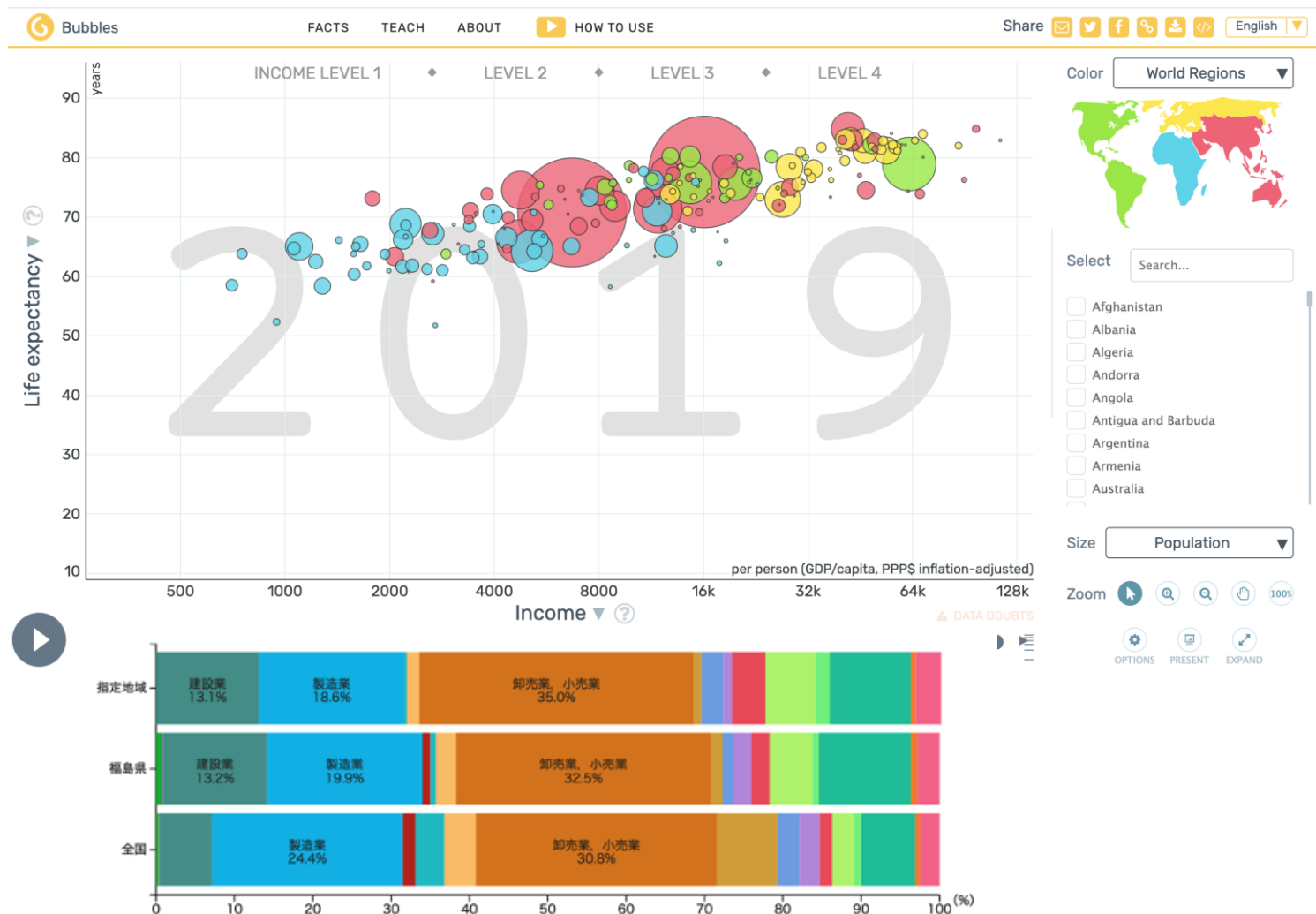
	A	B	C	D	E
1		2021年米の生産量（世界）			
2	Area Code (N	Area	Year	Unit	Value
3	159	China	2021	t	214403870
4	156	China, mainla	2021	t	212843000
5	356	India	2021	t	195425000
6	50	Bangladesh	2021	t	56944554
7	360	Indonesia	2021	t	54415294
8	704	Viet Nam	2021	t	43852728.7
9	764	Thailand	2021	t	33582000
10	104	Myanmar	2021	t	24910000
11	608	Philippines	2021	t	19960170.2
12	586	Pakistan	2021	t	13984009
13	76	Brazil	2021	t	11660603
14	116	Cambodia	2021	t	11410000
15	392	Japan	2021	t	10525400
16	840	United State	2021	t	8699720
17	566	Nigeria	2021	t	8342000
18	524	Nepal	2021	t	5621710
19	410	Republic of K	2021	t	5211037

定型データ

パネルデータ

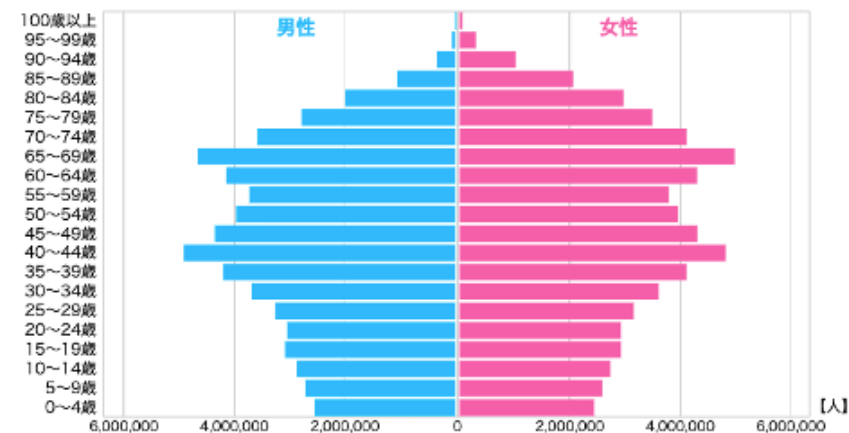
	A	B	C	D	E	F
1	米の生産量（日本とアメリカ）					
2	Area Code (N	Area	Item	Year	Unit	Value
3	392	Japan	Rice	2015	t	10925000
4	392	Japan	Rice	2016	t	10934000
5	392	Japan	Rice	2017	t	10777000
6	392	Japan	Rice	2018	t	10606000
7	392	Japan	Rice	2019	t	10540100
8	392	Japan	Rice	2020	t	10469000
9	392	Japan	Rice	2021	t	10525400
10	840	United States of America	Rice	2015	t	8724530
11	840	United States of America	Rice	2016	t	10167050
12	840	United States of America	Rice	2017	t	8084290
13	840	United States of America	Rice	2018	t	10152890
14	840	United States of America	Rice	2019	t	8396180
15	840	United States of America	Rice	2020	t	10319860
16	840	United States of America	Rice	2021	t	8699720

データの可視化



時点: 2015 年

全国 (日本)



データの種類

データを集める

データを読む

データを説明する

種類	事例	手段
調査データ	国勢調査・政府統計調査	アンケート
	民間の信用調査	
ログデータ	検索ログ	検索エンジン
	購買ログ	レシート
	人の行動ログ	スマートフォン
	車の移動ログ	カーナビ
	機械の稼働ログ	センサ
実験データ	研究データ	
観測データ	気象予測	AMeDAS センサ
	河川の水位データ	

オープンデータ



オープンデータの定義

2 次利用可能

- 営利・非営利にかかわらず利用できる

機械判読可能

- コンピュータで処理できる
- 加工・編集・再配布ができる

無償利用可能

https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/

オープンデータの定義

2次利用可能

- 営利・非営利にかかわらず利用できる

機械判読可能

- コンピュータで処理できる
- 加工・編集・再配布ができる

無償利用可能

(c) 平成 23 年 3 月 14 日 21 時 25 分頃以降の放射線量上昇との関係

① 正門付近における放射線量測定の結果、平成 23 年 3 月 14 日 21 時 25 分頃に $6.8 \mu\text{Sv/h}$ であったのに対し、同日 21 時 35 分頃に $760.0 \mu\text{Sv/h}$ 、同日 21 時 37 分頃に $3,130.0 \mu\text{Sv/h}$ に急上昇した（図表 16 参照）。

H23. 3. 14 時刻	放射線量 ($\mu\text{Sv/h}$)
21:15	5.8
21:20	6.0
21:25	6.8
21:30	29.7
21:35	760.0
21:37	3,130.0

② まず、これらの放射線量上昇は、1 号機及び 3 号機の R/B 爆発や格納容器ベントの実施に伴い放射線量が上昇してから相当時間が経過し、既に放射線量が下降した後の挙動であるから、これらの事象による影響以外に要因があると考えるのが自然である。

図表16 正門付近の放射線量測定結果
東京電力「福島第一原子力発電所構内での計測データ」（平成23年5月）を基に作成

PDF

```
年,月,A市,B市,C市,D町
2018,1,-4.5,-0.5,1.6,11.3
2018,2,-6.8,-2.1,0.4,8.4
2018,3,-2.4,1.9,3.8,13.5
2018,4,0.2,3.4,6.5,17.3
```

CSV(Comma Separated Values)

	A	B	C	D	E	F
1	年	月	A市	B市	C市	D町
2	2018	1	-4.5	-0.5	1.6	11.3
3	2018	2	-6.8	-2.1	0.4	8.4
4	2018	3	-2.4	1.9	3.8	13.5
5	2018	4	0.2	3.4	6.5	17.3

Excel

https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/

データの取得

e-Stat 統計で見る日本
政府統計の総合窓口

統計データを探す 統計データの活用 統計データの高度利用 統計関連情報 リンク集

●統計データを探す (政府統計の調査結果を探します)

すべての分野 組織

政府統計一覧の中から探します 17の統計分野から探します 統計を作成した府省等から探します

キーワード検索: 例: 国勢調査 検索

●統計データを活用する

グラフ 時系列表 地図 地域

主要指標をグラフで表示 (統計ダッシュボード) 主要指標を時系列表で表示 (統計ダッシュボード) 地図上に統計データを表示 都道府県、市区町村の主要データを表示

人口・世帯



主な調査

- ＞ 国勢調査
- ＞ 人口推計
- ＞ 人口動態調査
- ＞ 生命表
- ＞ 国民生活基礎調査

＞ すべて見る (21 調査)




データの取得

国勢調査

[詳細](#)[一覧形式で表示](#)

国勢調査は、日本に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査で、5年ごとに実施されます。国勢調査から得られる日本の人口や世帯の実態は、国や地方公共団体の行政において利用されることはもとより、民間企業や研究機関でも広く利用され、そのような利用を通じて国民生活に役立てられています。

国勢調査では、年齢別の人口、家族構成、働いている人や日本に住んでいる外国人などの結果を提供しています。

国勢調査	データベース	件数 更新日	フォーマット	件数 更新日	概要
時系列データ		62件 2022-03-31		226件 2022-03-31	
都道府県・市区町村別の主な結果				5件 2022-07-22	

国勢調査

[詳細](#)[一覧形式で表示](#)

国勢調査は、日本に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査で、5年ごとに実施されます。国勢調査から得られる日本の人口や世帯の実態は、国や地方公共団体の行政において利用されることはもとより、民間企業や研究機関でも広く利用され、そのような利用を通じて国民生活に役立てられています。

国勢調査では、年齢別の人口、家族構成、働いている人や日本に住んでいる外国人などの結果を提供しています。

時系列データ	公開（更新）日
■男女、年齢、配偶関係 [52件]	2022-03-31
■人口の労働力状態、就業者の産業・職業 [149件]	2017-12-18

データの取得

国勢調査

[詳細](#)[一覧形式で表示](#)

国勢調査は、日本に住んでいるすべての人と世帯を対象とする国の最も重要な統計調査で、5年ごとに実施されます。国勢調査から得られる日本の人口や世帯の実態は、国や地方公共団体の行政において利用されることはもとより、民間企業や研究機関でも広く利用され、そのような利用を通じて国民生活に役立てられています。

国勢調査では、年齢別の人口、家族構成、働いている人や日本に住んでいる外国人などの結果を提供しています。

時系列データ	公開（更新）日
■男女、年齢、配偶関係 [52件]	2022-03-31
■人口の労働力状態、就業者の産業・職業 [149件]	2017-12-18
■世帯 [21件]	2022-03-31
■従業地・通学地 [1件]	2017-07-06
■CSV形式による主要時系列データ [3件]	2019-05-14

データセット一覧

[戻る](#)[URLをコピー](#)[一覧形式で表示](#)

政府統計名	国勢調査	詳細
提供統計名	時系列データ	
提供分類1	CSV形式による主要時系列データ	

表番号	統計表	調査年月	公開（更新）日	ダウンロード
1	男女別人口ー全国、都道府県（大正9年～平成27年）	2015年	2017-01-30	CSV
2	年齢（5歳階級）、男女別人口ー全国（大正9年～平成27年）	2015年	2017-01-30	CSV
3	年齢（5歳階級）、男女別人口ー都道府県（大正9年～平成27年）	2015年	2019-05-14	CSV

✖ データ損失の可能性 このブックをコンマ区切り (.csv) 形式で保存すると、一部の機能が失われる可能性があります。機能が失われない。

A1									
	A	B	C	D	E	F	G	H	I
1	都道府県コード	都道府県名	元号	和暦（年）	西暦（年）	注	人口（総数）	人口（男）	人口（女）
2	0	全国	大正	9	1920		55963053	28044185	27918868
3	1	北海道	大正	9	1920		2359183	1244322	1114861
4	2	青森県	大正	9	1920		756454	381293	375161
5	3	岩手県	大正	9	1920		845540	421069	424471
6	4	宮城県	大正	9	1920		961768	485309	476459
7	5	秋田県	大正	9	1920		898537	453682	444855
8	6	山形県	大正	9	1920		968925	478328	490597
9	7	福島県	大正	9	1920		1362750	673525	689225
10	8	茨城県	大正	9	1920		1350400	662128	688272
11	9	栃木県	大正	9	1920		1046479	514255	532224
12	10	群馬県	大正	9	1920		1052610	514106	538504
13	11	埼玉県	大正	9	1920		1,141,122	571,133	569,989
14	12	千葉県	大正	9	1920		1,141,122	571,133	569,989
15	13	東京都	大正	9	1920		1,141,122	571,133	569,989
16	14	神奈川県	大正	9	1920		1,141,122	571,133	569,989
17	15	新潟県	大正	9	1920		1,141,122	571,133	569,989
18	16	富山県	大正	9	1920		1,141,122	571,133	569,989
19	17	石川県	大正	9	1920		1,141,122	571,133	569,989
20	18	福井県	大正	9	1920		1,141,122	571,133	569,989
21	19	山梨県	大正	9	1920		1,141,122	571,133	569,989
22	20	長野県	大正	9	1920		1,141,122	571,133	569,989

c01

Excel ブック (*.xlsx)

Excel ブック (*.xlsx)

Excel マクロ有効ブック (*.xlsm)

Excel バイナリブック (*.xlsb)

Excel 97-2003 ブック (*.xls)

CSV UTF-8 (コンマ区切り) (*.csv)

XML データ (*.xml)

単一ファイル Web ページ (*.mht, *.mhtml)

保存

データの取得

ファイル ホーム 挿入 描画 ページレイアウト 数式 **データ** 校閲 表示 開発 ヘルプ Acrobat

データの取得 テキストまたは CSV から Web から テーブルまたは範囲から 画像から 最近使ったソース 既存の接続 クエリと接続 クエリと接続 プロパティ リンクの編集 データの種類 株式 (Engli... 通貨 (Engli... 並べ替え フィルター クリア 再適用 詳細設定

A1 : \times \checkmark f_x 都道府県コード

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	都道府県	都道府県	元号	和暦 (年)	西暦 (年)	注	人口 (総)	人口 (男)	人口 (女)				
2	0 全国		大正		9	1920	55963053	28044185	27918868				
3	1 北海道		大正		9	1920	2359183	1244322	1114861				
4	2 青森県		大正		9	1920	756454	381293	375161				
5	3 岩手県		大正		9	1920	845540	421069	424471				

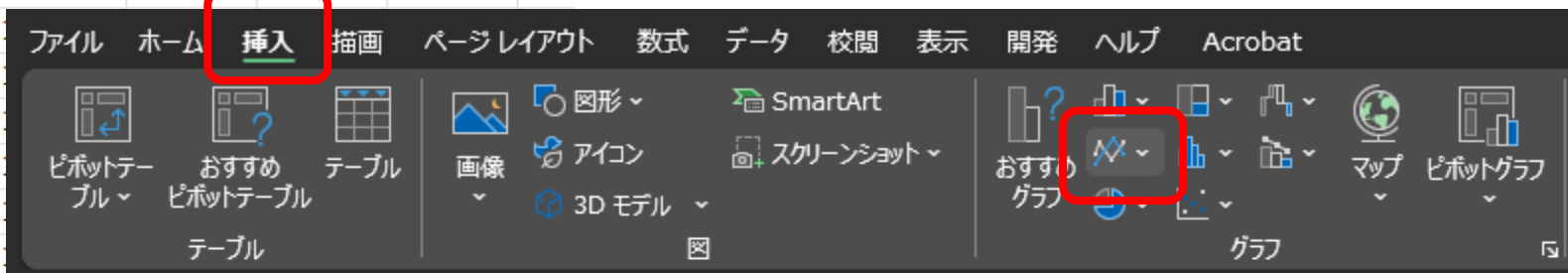
色で並べ替え(I) >
シートビュー(V) >
"都道府県コード" からフィルターをクリア(C)
色フィルター(I) >
数値フィルター(E) >
検索
☑ (すべて選択)
☑ 0
☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6
OK キャンセル

「0」のみチェックする

可視化

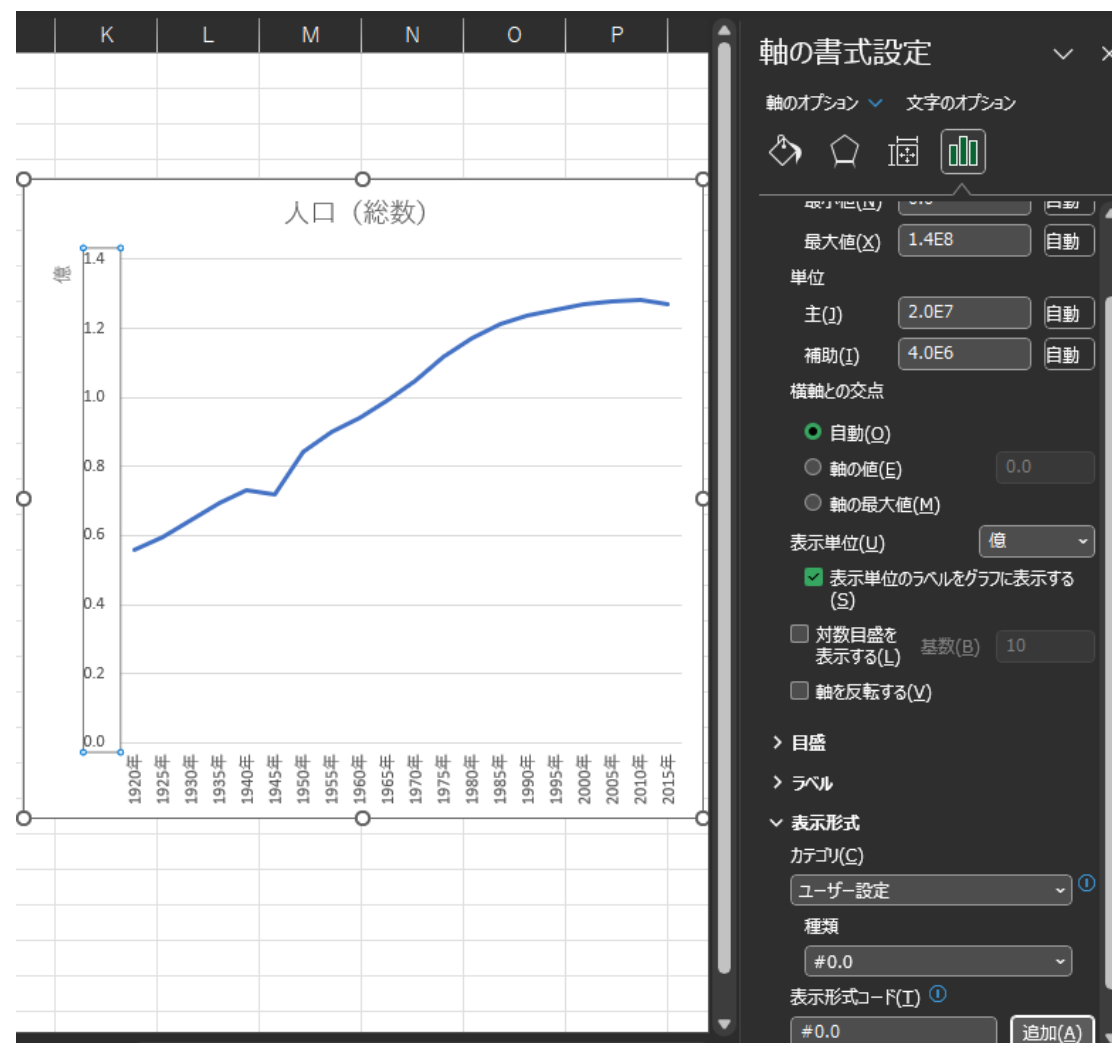
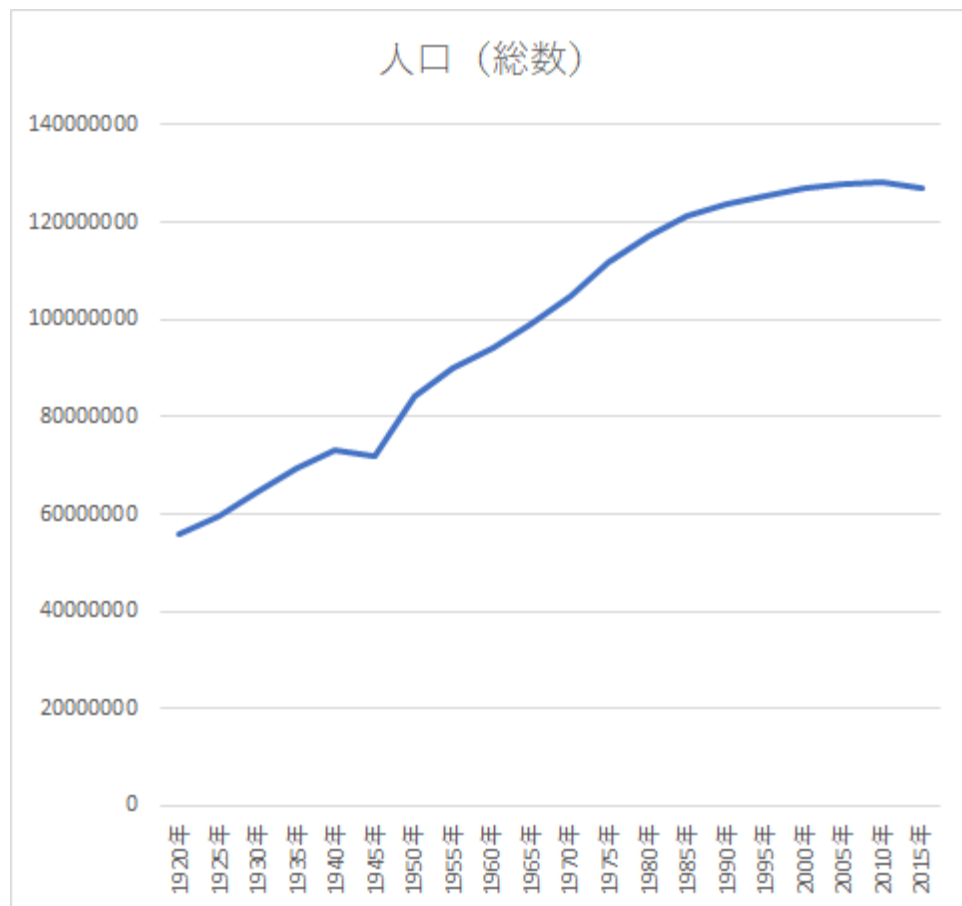
	A	B	C	D	E	F	G	H	I
1	都道府県	都道府県名	元号	和暦 (年)	西暦 (年)	西暦 (年)	人口 (総数)	人口 (男)	人口 (女)
2	0	全国	大正	9	1920	1920年	55963053	28044185	27918868
3	0	全国	大正	14	1925	1925年	59736822	30013109	29723713
4	0	全国	昭和	5	1930	1930年	64450005	32390155	32059850
5	0	全国	昭和		1935				
6	0	全国	昭和		1940				
7	0	全国	昭和		1945				
8					1950				
9					1955				
10					1960				
11	0	全国	昭和	40	1965				
12	0	全国	昭和	45	1970				
13	0	全国	昭和	50	1975				
14	0	全国	昭和	55	1980				
15	0	全国	昭和	60	1985				
16	0	全国	平成	2	1990				
17	0	全国	平成	7	1995				
18	0	全国	平成	12	2000				
19	0	全国	平成	17	2005				
20	0	全国	平成	22	2010				
21	0	全国	平成	27	2015				

=E2 & “年”



	A	B	C	D	E	F	G	H	I	J	K
1	都道府県	都道府県名	元号	和暦 (年)	西暦 (年)	西暦 (年)	人口 (総数)	人口 (男)	人口 (女)		
2	0	全国	大正	9	1920	1920年	55963053	28044185	27918868		
3	0	全国	大正	14	1925	1925年	59736822	30013109	29723713		
4	0	全国	昭和	5	1930	1930年	64450005	32390155	32059850		
5	0	全国	昭和	10	1935	1935年	69254148	34734133	34520015		
6	0	全国	昭和	15	1940	1940年	73114308	36566010	36548298		
7	0	全国	昭和	20	1945	1945年	71998104	33894059	38104045		
8	0	全国	昭和	25	1950	1950年	84114574	41241192	42873382		

可視化



可視化

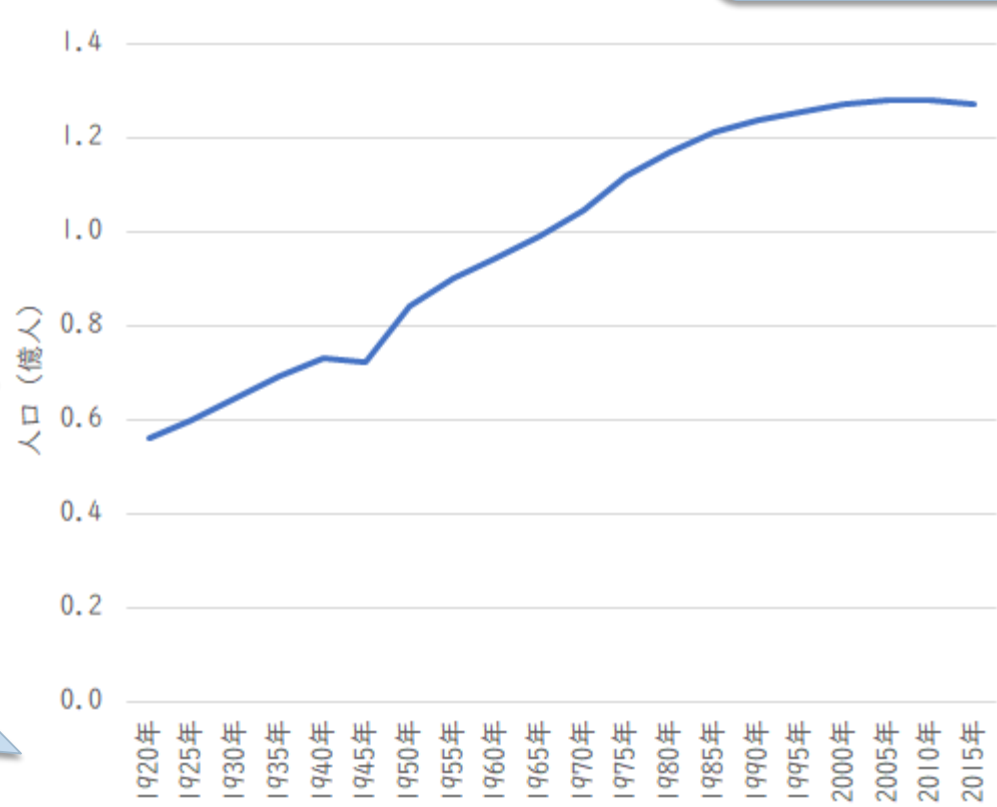
グラフタイトル

メイリオ

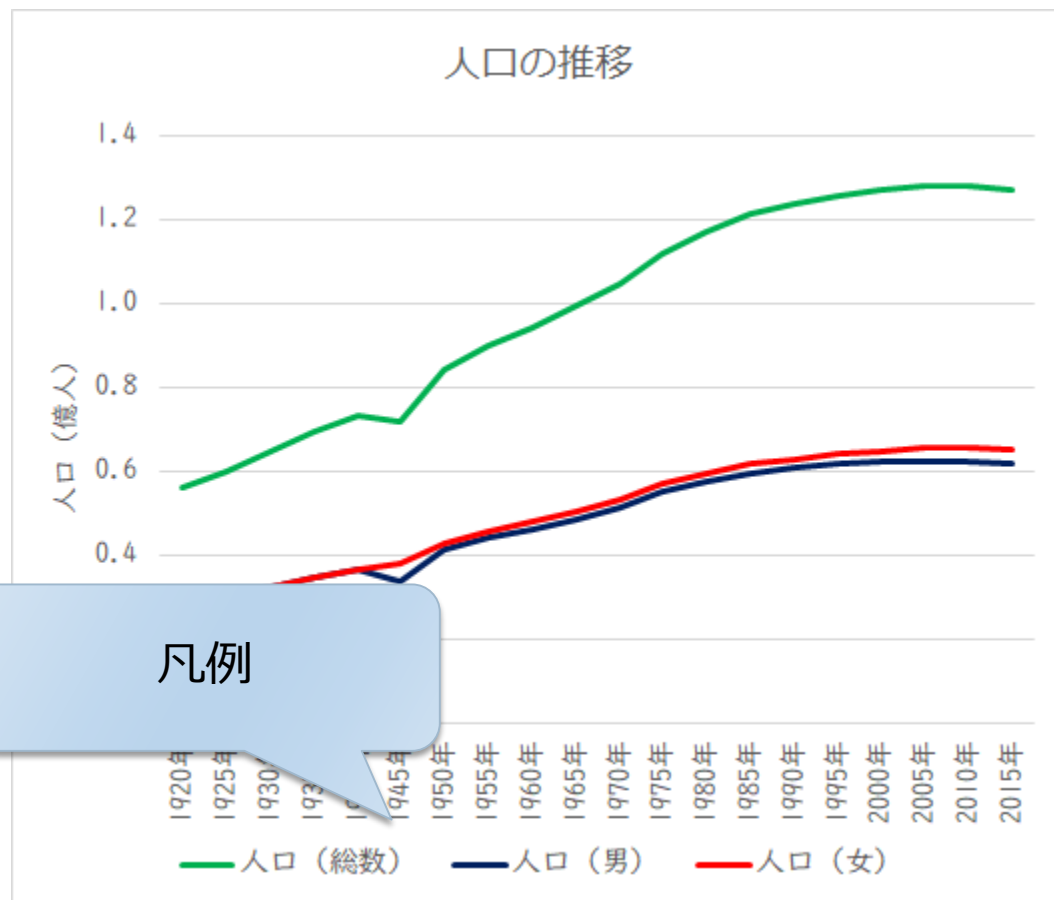
人口（総数）

軸ラベル
(単位)

ユニバーサル
デザインフォント



可視化



指数化（変化の大きさ）

基準年度の値を100として、他の年度の値を相対値として表したもの

- ▶ 全体での水準の大きさ情報を除く
- ▶ 変化のパターン
- ▶ 変化の大きさ
- ▶ 水準や単位の違う系列間での変化の比較

$$\text{指数}_t = \frac{data_t}{data_0} \times 100$$

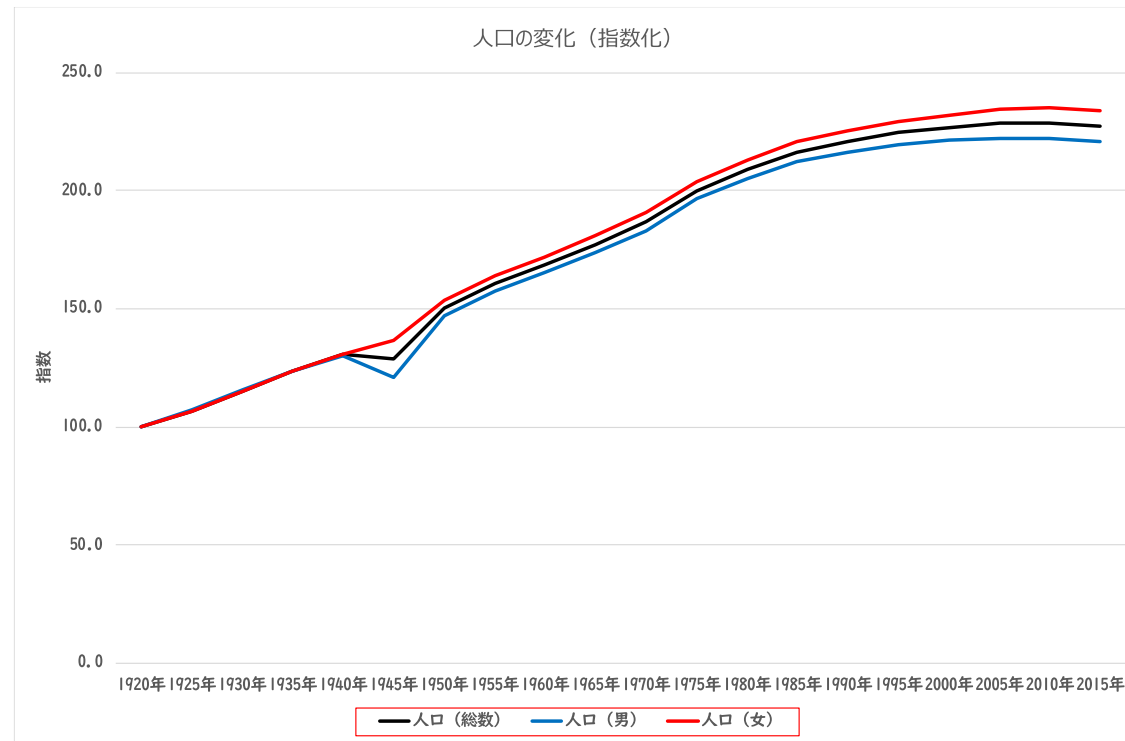
$data_t$: t時点のデータ

$data_0$: 基準値

指数化

F	G	H	I	J	K	L	M	N
西暦（年）	人口（総数）	人口（男）	人口（女）		西暦（年）	人口（総数）	人口（男）	人口（女）
1920年	55963053	28044185	27918868		1920年	100.0	100.0	100.0
1925年	59736822	30013109	29723713		1925年	106.7	107.0	106.5
1930年	64450005	32390155	32059850		1930年	115.2	115.5	114.8
1935年	69254148	34734133	34520015		1935年	123.7	123.9	123.6
1940年	73114308	36566010	36548298		1940年	130.6	130.4	130.9
1945年	71998104	33894059	38104045		1945年	128.7	120.9	136.5
1950年	84114574	41241192	42873382		1950年	150.3	147.1	153.6
1955年	90076594	44242657	45833937		1955年	161.0	157.8	164.2
1960年	94301623	46300445	48001178		1960年	168.5	165.1	171.9
1965年	99209137	48692138	50516999		1965年	177.3	173.6	180.9
1970年	1.05E+08	51369177	53295994		1970年	187.0	183.2	190.9
1975年	1.12E+08	55090673	56848970		1975年	200.0	196.4	203.6
1980年	1.17E+08	57593769	59466627		1980年	209.2	205.4	213.0
1985年	1.21E+08	59497316	61551607		1985年	216.3	212.2	220.5
1990年	1.24E+08	60696724	62914443		1990年	220.9	216.4	225.3
1995年	1.26E+08	61574398	63995848		1995年	224.4	219.6	229.2
2000年	1.27E+08	62110764	64815079		2000年	226.8	221.5	232.2
2005年	1.28E+08	62348977	65419017		2005年	228.3	222.3	234.3
2010年	1.28E+08	62327737	65729615		2010年	228.8	222.2	235.4
2015年	1.27E+08	61841738	65253007		2015年	227.1	220.5	233.7

$$=I2/I\$2*100$$



fn キー



Windowsのノートパソコンにも、Fnキーを押し
ながらF4キーを押す必要がある機種もあります。



成長率（変化の勢い）

- ▶ 1 時点前からの変化の大きさの推移をみる
 - ▶ 季節変動がある場合には、対前年同月比

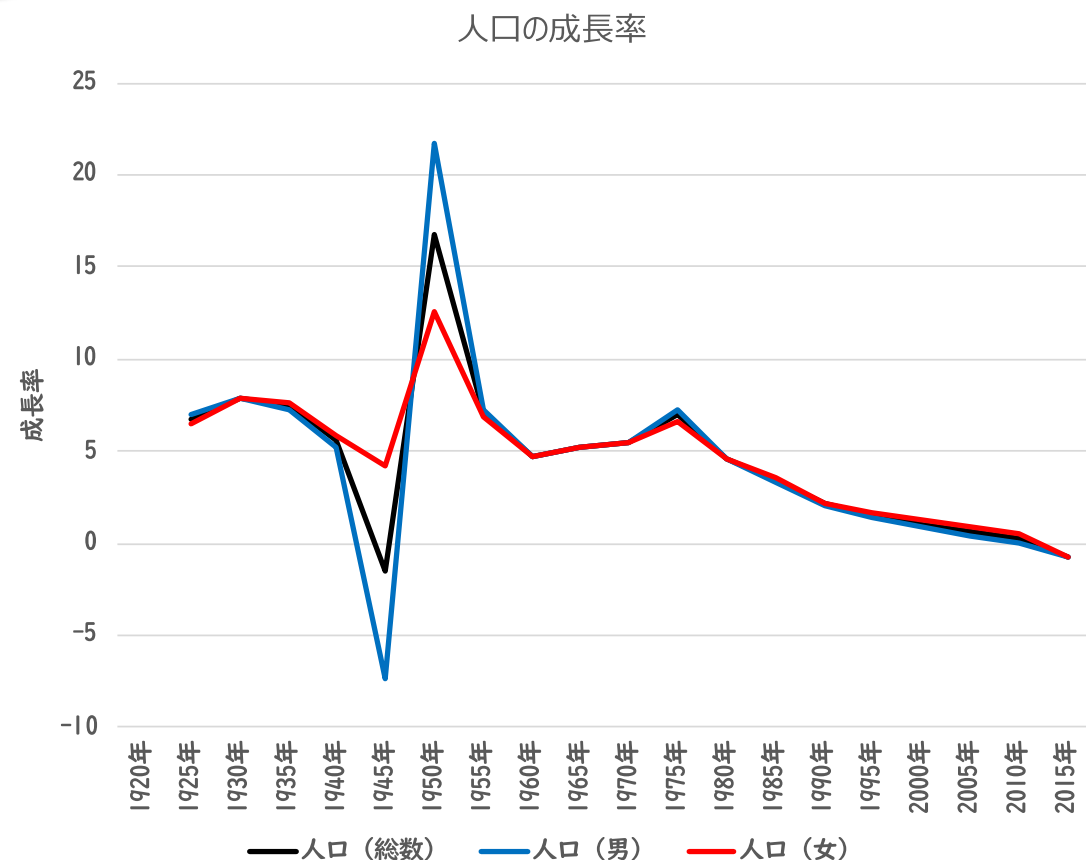
$$\text{成長率}_t = \frac{data_t - data_{t-1}}{data_{t-1}} \times 100$$

$data_t$: t時点のデータ
 $data_{t-1}$: tに対し、1 時点前のデータ

成長率

F	G	H	I	O	P	Q	R	S
西暦（年）	人口（総数）	人口（男）	人口（女）		西暦（年）	人口（総数）	人口（男）	人口（女）
1920年	55963053	28044185	27918868		1920年			
1925年	59736822	30013109	29723713		1925年	6.7433222	7.020792	6.464607
1930年	64450005	32390155	32059850		1930年	7.8899125	7.920026	7.859506
1935年	69254148	34734133	34520015		1935年	7.4540615	7.236699	7.673663
1940年	73114308	36566010	36548298		1940年	5.5739044	5.273997	5.875672
1945年	71998104	33894059	38104045		1945年	-1.526656	-7.3072	4.256688
1950年	84114574	41241192	42873382		1950年	16.828874	21.67676	12.51662
1955年	90076594	44242657	45833937		1955年	7.087975	7.277833	6.905345
1960年	94301623	46300445	48001178		1960年	4.6904849	4.65114	4.728464
1965年	99209137	48692138	50516999		1965年	5.204061	5.165594	5.241165
1970年	1.05E+08	51369177	53295994		1970年	5.4995277	5.497888	5.501109
1975年	1.12E+08	55090673	56848970		1975年	6.9502318	7.244609	6.666497
1980年	1.17E+08	57593769	59466627		1980年	4.5745661	4.543593	4.604581
1985年	1.21E+08	59497316	61551607		1985年	3.4072386	3.305127	3.506135
1990年	1.24E+08	60696724	62914443		1990年	2.1167012	2.015903	2.214136
1995年	1.26E+08	61574398	63995848		1995年	1.5848722	1.445999	1.71885
2000年	1.27E+08	62110764	64815079		2000年	1.0795527	0.871086	1.280131
2005年	1.28E+08	62348977	65419017		2005年	0.6634984	0.383529	0.931786
2010年	1.28E+08	62327737	65729615		2010年	0.2264714	-0.03407	0.474782
2015年	1.27E+08	61841738	65253007		2015年	-0.7517	-0.77975	-0.7251

$$= (I3 - I2) / I2 * 100$$



まとめ

項目	内容		
データの分類	データ	定型データ	時系列データ・クロスセクションデータ・パネルデータ
		非定型データ	画像・音声・文章
オープンデータの定義	二次利用可能	ビジネスに利用可能	
	機会判読可能	CSV・Excel > PDF	
	無償利用可能		
CSV	テキスト Excelで利用可能		
時系列データの指数化	基準年度の値を100として、他の年度の値 を相対値として表したもの		
時系列データの成長率	1 時点前からの変化の大きさの推移をみる（勢いを表す）		
Excel	式の入れ方・絶対参照 グラフの描き方		

参考文献

- ▶ 「身近な統計」 石崎克也・渡辺美智子 放送大学教育振興会
- ▶ 「誰でも使える統計オープンデータ」 オフィシャル スタディノート
総務省統計局