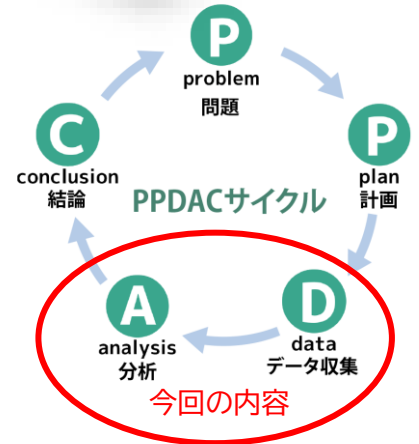




1. はじめに

統計を活用した探究活動に「PPDAC サイクル」というのがあります。問題提起 (Problem), 計画 (Plan), データ収集 (Data), 分析 (Analysis), 結論 (Conclusion) です。今回は主にインターネット上にあるデータを使って「データ収集 (Data), 分析 (Analysis)」について紹介します。

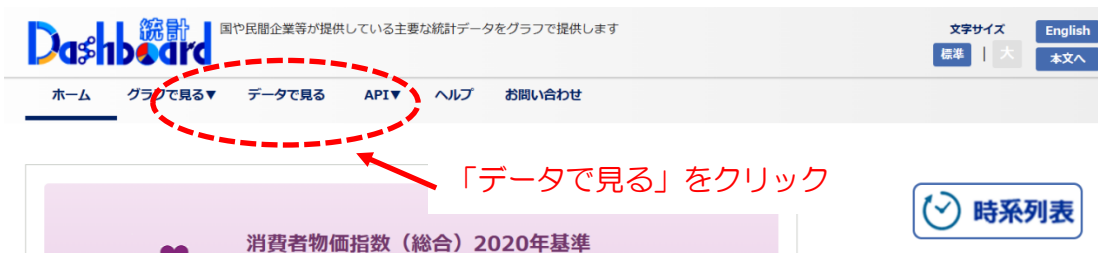


2. データ収集のポイント

インターネットを活用すれば簡単に多くのデータを収集することができます。しかし、インターネットは簡単にデータをアップできるため、作成者にデータについての責任が乏しいため信頼できないデータも多くあります。

3. データ収集の方法例

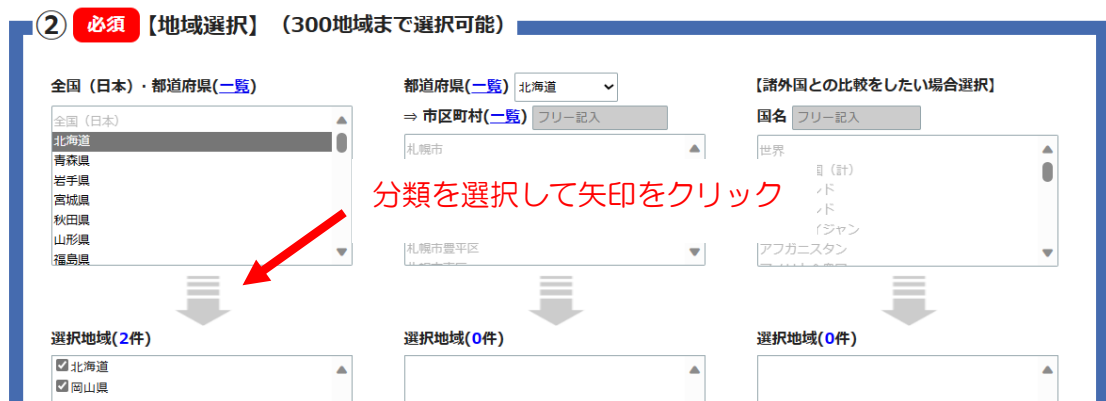
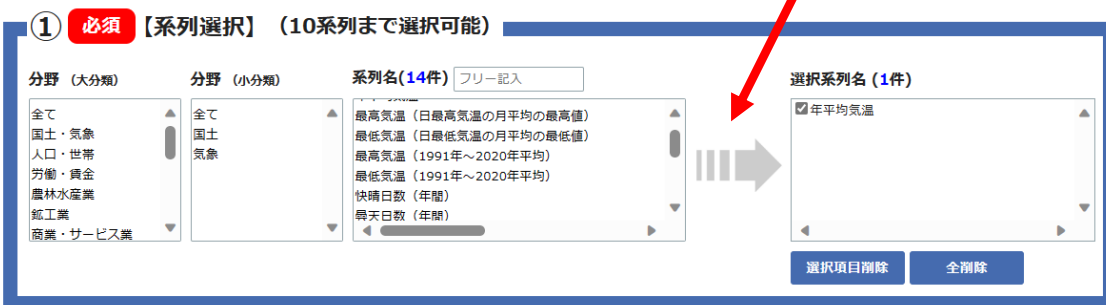
今回はデータの出典先が信頼できる「統計 Dashboard」を使ってデータ分析をしてみよう。まず、アドレス 「 <https://dashboard.e-stat.go.jp/> 」を入力します。



【例】北海道と岡山の年平均気温データを取り出そう

以下の必須項目を選択します。

分類を選択して矢印をクリック



③【時間選択】

データ周期: 全て 月 四半期 年 年度

時点: 全て 指定: 1974 年から 2023 年まで

原数値・季節調整値: 全て 原数値 季節調整値

期間を指定して、表示をクリック

検索結果(94件)

※最大1000件まで表示可。最大100万件までダウンロード可

月 (0)		
▲時点	北海道	岡山県
1975年	8.3	14.9
1976年	7.9	14.1
1977年	7.8	14.9
1978年	8.2	15.2
1979年	8.4	15.4
1980年	7.9	14.1
1981年	7.8	14.2
1982年	8.9	-
1983年	8.3	16
1984年	8.1	15.3
1985年	8.3	15.8
1986年	7.7	15.2
1987年	8.2	16.1
1988年	8.2	15.3
1989年	9.4	16
1990年	10.1	16.8
1991年	9.5	16.1
1992年	8.8	15.8

表示を押すと、右のように北海道と岡山の年平均気温のデータが表示されます。


4. データを分析してみよう

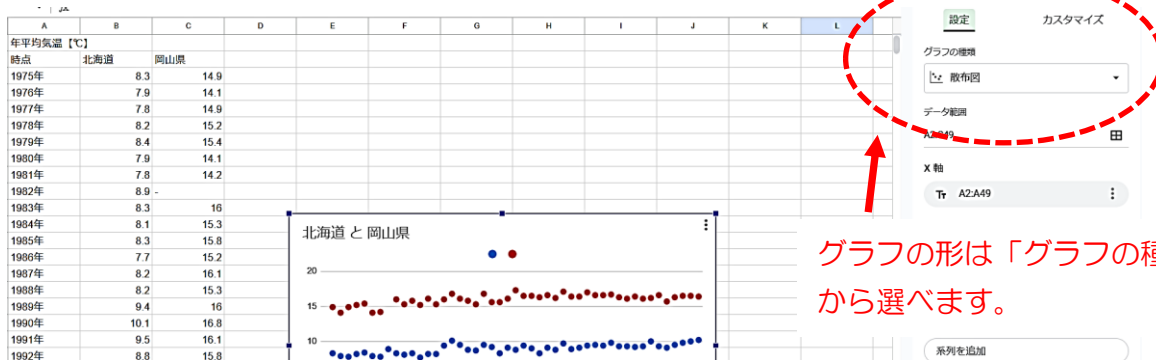
次に、このデータを使って以下の問を考えてみよう。

「北海道と岡山では 1975 年から 2021 年までの年平均気温の上昇度合いはどちらが大きいのか？」

【仮説】北海道の方が年平均気温の上昇度合いはどちらが大きい

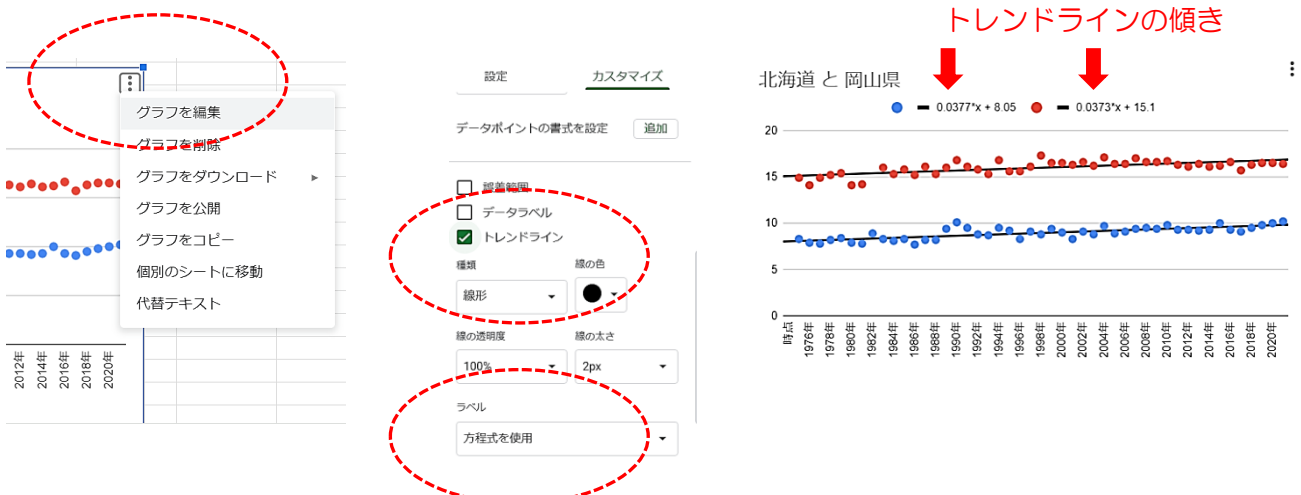
分析1：グラフにして考えてみよう。

- ① 齊木ほどの表の画面にあるダウンロードボタンを押します。その後データ一覧の画面で左上の「ファイル」→「Google スプレッドシートとして保存」を選択します。
- ② データを範囲指定してグラフボタン  を押します。



グラフの形は「グラフの種類」から選べます。

さらにできたグラフを選択し、右上のドットをクリックすると、「グラフを編集」からグラフの加工ができます。そして、「カスタマイズ→系列→トレンドライン」を選ぶと、データ変化の傾向を示す「近似直線」を引くことができます。最後に、「ラベル」で「方程式を利用」を選ぶと、グラフの上に、直線の方程式が出ます。この傾きで年平均気温の変化の大きさが比較できます。



分析の結果、2直線のグラフの傾きにあまり違いは見られませんでした。よって、「北海道と岡山の1975年から2021年までの年平均気温の上昇度合いに違いは見られない」という結論になります。

分析2：基本統計量を算出して考えてみよう。

「統計」とは集団を数量的に把握することです。「**基本統計量**」とは平均値、合計値、データ数、最大値、最小値、分散、標準偏差など色々あります。これらを算出してデータの特徴を推測します。先ほどの北海道と岡山の平均気温のデータについて、右の関数を用いて基本統計量を求めて比較してみよう。どこかのセルに関数を入力してみましよう。○の部分にはセル番地が入ります。【例】A1 など

合計	=SUM(○ : ○)
データ数	=COUNT(○ : ○)
平均	=AVERAGE(○ : ○)
分散	=VARP(○ : ○)
標準偏差	=STDEV(○ : ○)
最小値	=MIN(○ : ○)
第1四分位数	=QUARTILE(○ : ○ , 1)
第2四分位数(中央値)	=QUARTILE(○ : ○ , 2)
第3四分位数	=QUARTILE(○ : ○ , 3)
最大値	=MAX(○ : ○)

例えば A1 から B10 までのデータの「平均」を表す関数は「 **=AVERAGE(A1:B10)** 」です。ちなみに、北海道と岡山の平均気温について平均を求めると、それぞれ 9.0°C、16°Cです。これらの関数は覚えておくと便利です。

分析3（発展）：いろいろな統計手法を用いて分析してみよう。

今回のように2種類のデータの基本統計量を算出してみて、例えばデータのちらばり（分散）に違いがあると言えるかどうか、その1つの指標となる統計手法として「分散の検定（通称 F 検定）」という方法があります。の関数は「 **=FTEST(○ : ○, ○ : ○)** 」です。色々な統計手法について知りたい人は【付録2】を読んで自分で調べてみましょう。

5. 最後に

データを分析する前には必ず、「仮説（予想）」をたてましょう。 数ある分析方法の中からどの分析方法が最も適しているのかは自分で考えなければいけません。適当に分析してみてそこから仮説を考えるのではなく、はじめに何に着目して分析するのか仮説を決めてから始めましょう。

それから統計分析の結果はあくまでも取得したデータ内での話です。よって「～の傾向がみられる。」「ほぼ～と言える。」といった結論が一般的です。断言するためにはよほど信頼できるデータで、様々な分析方法を用いて明確な差がある場合に限ります。くれぐれも分析方法や統計手法では 100%言い切れないことに注意しましょう。

今回の基礎統計学で用いた方法は、「課題研究」の分析でも活用できます。各グループのデータに潜む規則性や法則、傾向の発見を目指してください。

6. 自分でデータを分析してみよう【重要】

統計 Dashboard では他にも色々なデータを取得することができます。自分で選んでやってみよう。

- 各クラスルームの「授業」にある課題「R05_基礎統計学ワークシート」を開きましょう。
- そして、統計 Dashboard で入手したデータをコピーして貼り付けましょう。
- 分析する前に「分析テーマ、使用するデータ、仮説」を入力しましょう。
- 分析ができれば分析結果を班で発表しあおう。最後に提出ボタンを押して提出しよう。



【付録1】引用・参考HP

- Data Start 総務省統計局 <https://www.stat.go.jp/dstart/>
- JDSSP 高等学校データサイエンス教育研究会
<https://ds-education.com/data-science-text/174/>
- DATA GO.JP（日本政府全体のデータ一覧サイト）
- e-Stat（政府統計の総合窓口）
- 統計GIS（地図で見る統計）
- RESAS（リーサス）地域経済分析システム（データがすぐに可視化できるサイト）
- SSDSE（教育用標準データセット）統計データ分析
- グローバルノート（様々なオープンデータがあります。）
- miripo（マイクロデータ利用ポータルサイト）
- 統計ダッシュボード（総務省統計局）
- なるほど統計学園
- 国土交通省 国土地理院

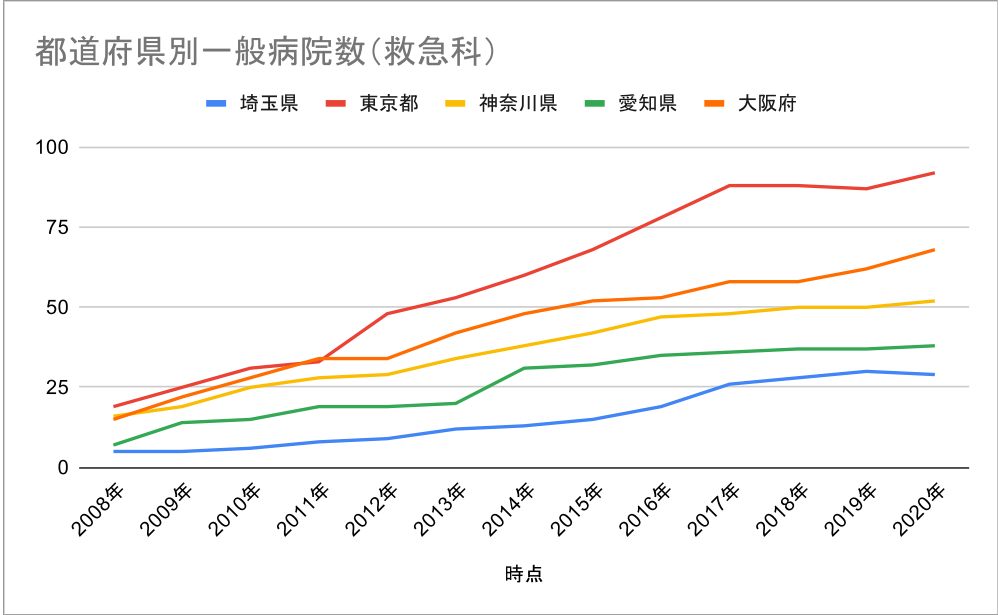
【付録2】統計手法を用いたデータ分析について

データを分析する際には、統計分析を使って予想（仮説）が成立する（もしくは成立しない）証拠を集めます。以下に統計的手法の例を紹介します。

- 原因を分析する ⇒ 重回帰分析
- 同じものを集めて共通性を分析する ⇒ デンドログラム（樹形図）、クラスター分析
- 特徴をまとめる ⇒ 主成分分析
- 表面上に出てこない要因を探す ⇒ 因子分析
- 仮説が正しいかどうかを判定する ⇒ t 検定 F 検定 χ^2 検定など
- 互いの関連性を判定する ⇒ 相関係数

これらの統計手法については google のスプレッドシートや Microsoft のアプリケーションソフトの Excel を使うと簡単にできます。

6	組	番	氏名
使用するデータ	人口の多い5都府県の 2008年から2020年までの一般病院数(救急科)		
仮説	人口が多いほど一般病院数の数は多く、2008年から2020年にかけてそれぞれ増加している。		
結論	救急科がある病院の数が 2008年から2020年にかけて増加していることは仮説どおりであった。また仮説通りであれば人口順で見ると救急科がある病院の数は東京都、神奈川県、大阪府、愛知県、埼玉県の順で多いという結果になる。しかしグラフを見てみると実際には 2009年以降に神奈川県 -大阪府間で病院数に大きく違いが見られた。そこで人口密度が大きいほど病院数が多いという仮説を改めて立ててグラフを比較してみると、埼玉県 -愛知県間の病院数が 2008年から2020年にかけて仮説との違いが生じているためこれも仮説とは異なる結果となった。		

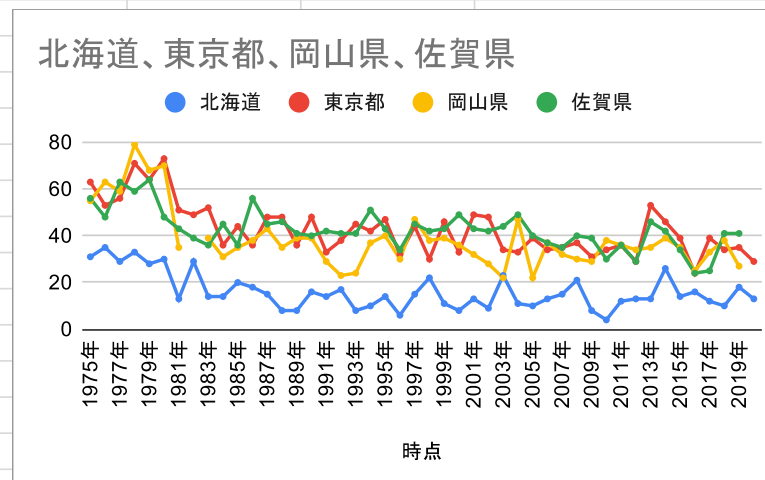


東京都	14,047,594人	
神奈川県	9,237,337人	
大阪府	8,837,685人	
愛知県	7,542,415人	
埼玉県	7,344,765人	

#	都道府県	人口密度
1	東京都	6,399.46
2	大阪府	4,611.99
3	神奈川県	3,821.01
4	埼玉県	1,931.98
5	愛知県	1,449.29

3	組		番	氏名	
分析テーマ	岡山県が「晴れの国」なのは本当か				
使用するデータ	1975年から2020年における年間の快晴日数				
仮説	岡山県は「晴れの国」であるため全国的に見ても快晴日数が多い。				
結論	快晴日数は岡山県が特別多いという傾向は見られなかった。→岡山県は快晴日数が多いために「晴れの国」なのではない。				

快晴日数(年間)【日】	北海道	東京都	岡山県	佐賀県
1975年	31	63	55	56
1976年	35	53	63	48
1977年	29	56	59	63
1978年	33	71	79	59
1979年	28	64	68	64
1980年	30	73	70	48
1981年	13	51	35	43
1982年	29	49	-	39
1983年	14	52	39	36
1984年	14	36	31	45
1985年	20	44	35	36
1986年	18	36	38	56
1987年	15	48	43	45
1988年	8	48	35	46
1989年	8	36	39	41
1990年	16	48	39	40
1991年	14	33	29	42
1992年	17	38	23	41
1993年	8	45	24	41
1994年	10	42	37	51

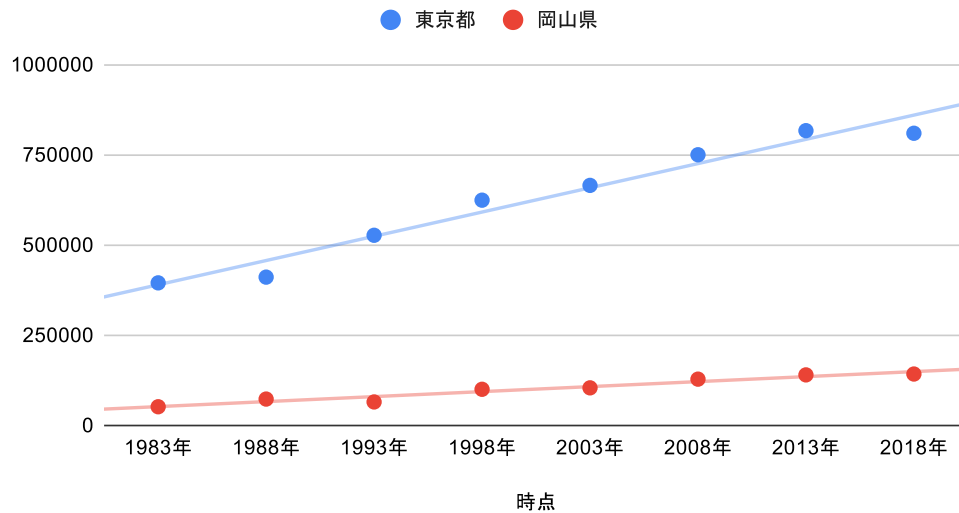


5	組		番	氏名	
分析テーマ	空き家件数				
使用するデータ	東京都、岡山県それぞれの空き家の軒数				
仮説	岡山県のほうが空き家が多い。理由：地方から都会に移住する人が増えており、岡山県でも東京に移住する人が増え、空き家が増えると考えた。また、東京都にはたくさんの移住者がいるので、岡山県よりも空き家が少ないと考えたから。				
結論	東京都のほうが多い。理由：日本の税制では、更地にするよりも何らかの建物がある方が固定資産税が安く、土地が売れず空き家を壊したり新しい建物を作る資金がない場合、空き家が放置されることが多い。また、東京では住んでいた人が亡くなって、その家や土地を相続する手続きがなされず空き家のまま放置されているケースがある。更に、分譲マンションなどで所有者が複数人いる場合は、建て壊しに所有者全員の許可が必要であるため、所有者の一人でも連絡がつかなければ、取り壊しができないという現状がある。				

空き家数【戸】

時点	東京都	岡山県
1983年	395200	51800
1988年	411100	73100
1993年	527100	65100
1998年	624400	100200
2003年	665400	104200
2008年	750300	128300
2013年	817100	140100
2018年	809900	142500

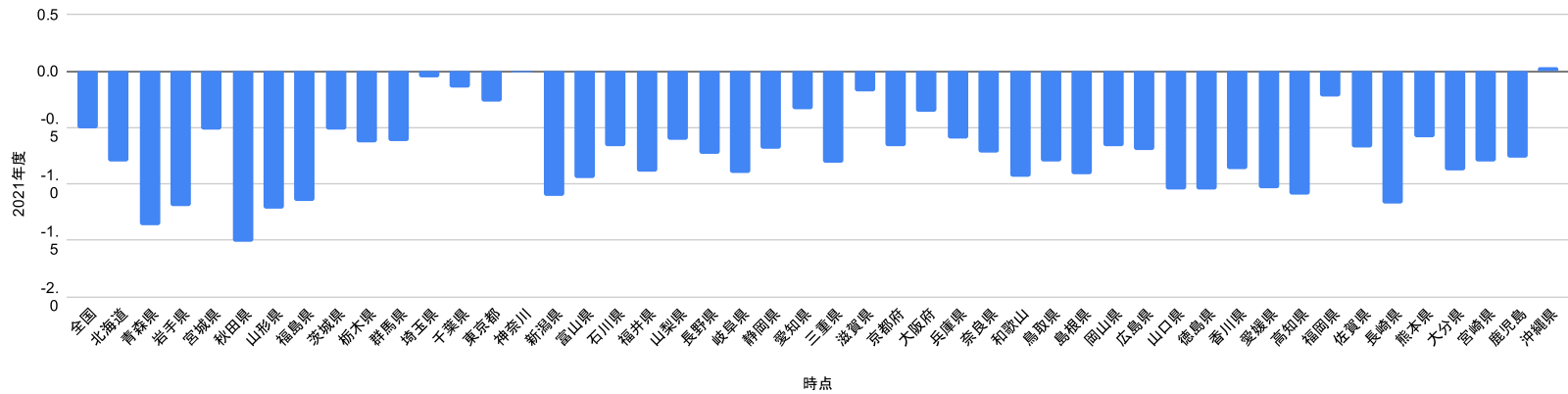
東京都と岡山県



3	組	番	氏名	考察:なぜ沖縄県だけ人口が増えているのか?		結果:おそらく自然増加	
分析テーマ	最も人口増加率が高い県			出生数【人】	転入者数【人】	死亡数【人】	転出者数【人】
使用するデータ	2021年度の各47都道府県の人口増減率			時点	沖縄県	沖縄県	沖縄県
仮説	移住人口が一番多そうな東京都			2021年	14,535	26,844	13,482
結論	沖縄県が人口増減率 0.04%で最も人口増減率が高い県だよ。(沖縄県以外は人口が減っている。)			出生数+転入者数 > 死亡数+転出者数		14535+26844=41379>40633=1	

人口増減率【%】	全国	北海道	青森県	岩手県	宮城県	秋田県	山形県	福島県	茨城県	栃木県	群馬県	埼玉県	千葉県
2021年度	-0.51	-0.8	-1.37	-1.2	-0.52	-1.51	-1.22	-1.15	-0.52	-0.63	-0.62	-0.06	-0.15

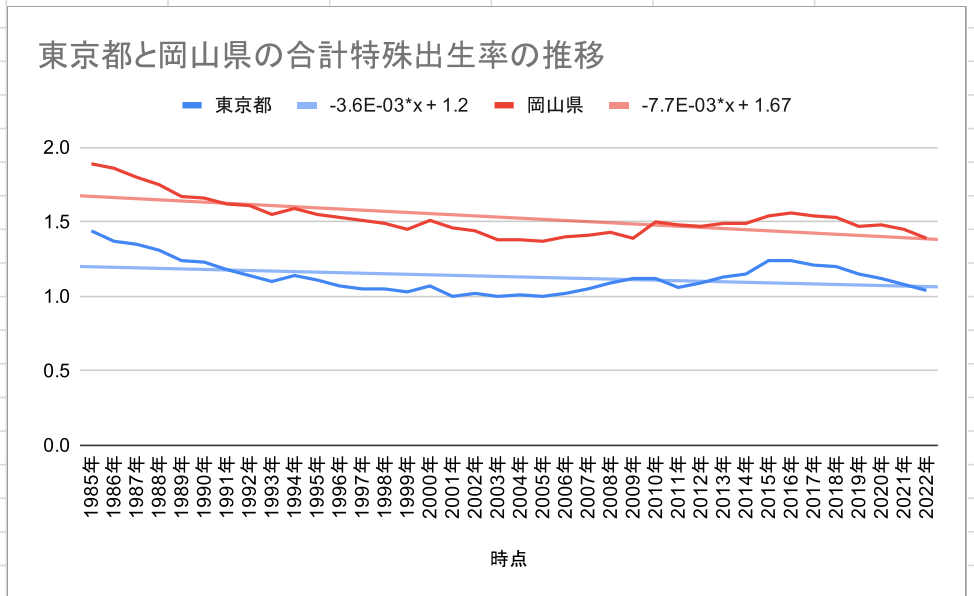
2021年度と時点



最小値: 秋田県の-1.51%
 最大値: 沖縄県の0.04%

5	組		番	氏名		
分析テーマ	東京都と岡山県の合計特殊出生率はどのように変わっているのか。					
使用するデータ	二地点の1985年から2022年までの合計特殊出生率					
仮説	東京都の方が、子育てのために住んでいる人は岡山県より少ないと予想されるため、合計特殊出生率は低くなる。					
結論	東京都の方が岡山県より合計特殊出生率が全体的に低いが、1985年から2022年までの差としては岡山県の方が大きい。					

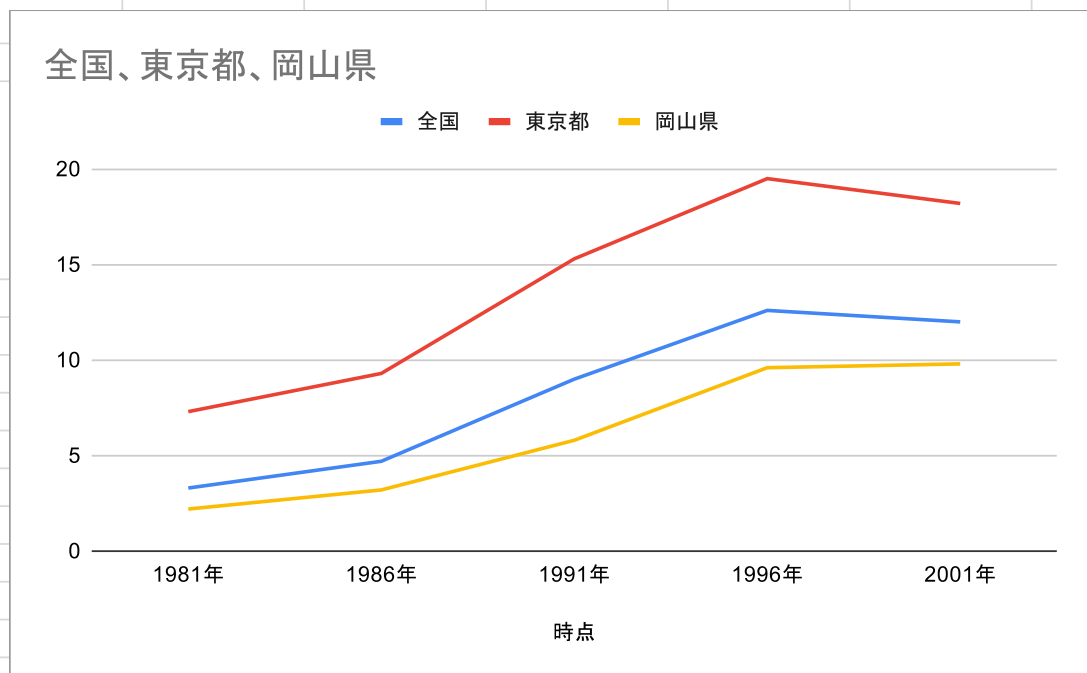
合計特殊出生率		
時点	東京都	岡山県
1985年	1.44	1.89
1986年	1.37	1.86
1987年	1.35	1.8
1988年	1.31	1.75
1989年	1.24	1.67
1990年	1.23	1.66
1991年	1.18	1.62
1992年	1.14	1.61
1993年	1.1	1.55
1994年	1.14	1.59
1995年	1.11	1.55
1996年	1.07	1.53
1997年	1.05	1.51
1998年	1.05	1.49
1999年	1.03	1.45
2000年	1.07	1.51
2001年	1	1.46
2002年	1.02	1.44
2003年	1	1.38
2004年	1.01	1.38
2005年	1	1.37



	東京都	岡山県
分散	0.0116166205	0.01610090028
最大値	1.44	1.89
最小値	1	1.37
差	0.44	0.52

1	組	番	氏名
分析テーマ	東京と岡山で海外旅行行動者率には差が出るのか？		
使用するデータ	海外旅行行動者率(15歳以上)【%】全国・東京・岡山		
仮説	いくら都会と田舎と言えど海外旅行に行く人の量は変わらないと思う。理由は住んでいる場所が違うだけで収入や行動力はバラバラだから！		
結論	岡山の人が東京はともかく、全国の行動率を下回っているのに衝撃を受けた。東京の海外旅行行動者率が全国を大きく上回る結果となりびっくりした。なぜ東京の人はこんなに海外旅行に行っているのかが気になる。行動力には違いがあるのかもしれない。		

海外旅行行動者率(15歳以上)【%】
 時点全国東京都岡山県
 1981年3.37.32.2
 1986年4.79.33.2
 1991年9.15.35.8
 1996年12.619.59.6
 2001年12.18.29.8



1	組	番	氏名						
分析テーマ	就業者と国内総生産の関係								
使用するデータ	就業者(男女計)【千人】、国内総生産(支出側)(名目)(米ドル表示)【米ドル】								
仮説	就業者が多い国でも、先進国でなければ国内総生産は少ないままである。反対に、就業者が少ない国でも技術で補い高い国内総生産になる国も存在すると考える。								
結論	中国、日本、アメリカが顕著な例で、中国では就業者に対する国内総生産が低く、日本、アメリカでは就業者に対する国内総生産が高い。								

就業者(男女計)【千人】									
時点	アルゼンチン	オーストラリア	オーストリア	ベルギー
1998年	8279	8517	3626	3857
1999年	8285	8673	3676	3987
2000年	8262	8902	3682	4120
2001年	8143	9015	3697	4039
2002年	8016	9190	3667	4052
2003年	8956	9397	3745	4055
2004年	9255	9558	3606	4144
2005年	9526	9881	3747	4235
2006年	9926	10124	3826	4264
2007年	10116	10434	3924	4380
2008年	10280	10731	3994	4446
2009年	10402	10806	3982	4421
2010年	10532	11022	4017	4489
2011年	10766	11214	4052	4509
2012年	10844	11351	4084	4524
2013年	10943	11457	4105	4530
2014年	11047	11540	4113	4544
2015年	...	11766	4148	4552
2016年	...	11973	4220	4586
2017年	11568	12252	4260	4638
2018年	11745	12584	4319	4755
2019年	12041	12875	4355	4832
2020年	10937	12675	4297	4802
2021年	12242	13065	4306	4854

国内総生産(支出側)(名目)(米ドル表示)【米ドル】									
時点	オーストラリア	ブラジル	中国	フランス
1995年	391094358364	778052821653	734484967620	1600996702405
1996年	434546469497	849622314042	863748958856	1605764901947

