

下山 紗代子

一般社団法人リンクデータ
代表理事

2020.11.18 (Wed.)

令和2年度
政策評価に関する統一研修

行政プロセスに
データ分析を
取り入れるために
知っておきたい
知識と事例



データ戦略 タスクフォースを はじめ、 国や自治体に おけるデータ活用 の検討に参加

https://www.kantei.go.jp/jp/singi/it2/dgov/data_strategy_tf/dai1/gijisidai.html

データ戦略タスクフォースの開催について

〔令和2年10月12日〕
デジタル・ガバメント関係会議決定

- 1 デジタル社会においては、データを十分に活用することが不可欠であり、競争力の源泉である。今般のコロナ危機では、迅速で的確な対応ができず、データ活用基盤（デジタルデータの整備、標準化、データの取扱いルール等）が官民ともに不十分であることが露呈した。デジタル国家にふさわしいデータ戦略を策定するためデータ戦略タスクフォース（以下「タスクフォース」という。）を開催する。
- 2 タスクフォースの構成は、次のとおりとする。ただし、主査は、必要があると認めるときは、構成員以外の関係行政機関の職員、有識者その他の関係者の出席を求めることができる。

主 査 内閣総理大臣補佐官（国土強靱化及び復興等の社会資本整備、地方創生、健康・医療に関する成長戦略並びに科学技術イノベーション政策その他特命事項担当）

構 成 員

砂金 信一郎	LINE 株式会社執行役員 AIカンパニーカンパニーCEO
遠藤 信博	一般社団法人日本経済団体連合会審議員会副議長 兼サイバーセキュリティ委員長 日本電気株式会社取締役会長
太田 直樹	株式会社 New Stories 代表取締役
小木曾 稔	一般社団法人新経済連盟政策部長
越塚 登	東京大学大学院教授
後藤 厚安	情報セキュリティ大学院大学学長
下山 紗代子	一般社団法人リンクデータ代表理事/インフォ・ラウンジ株式会社取締役
辻 昌彦	武蔵大学教授
手塚 悟	慶應義塾大学教授
村井 純	慶應義塾大学教授
渡部 俊也	東京大学未来ビジョン研究センター教授

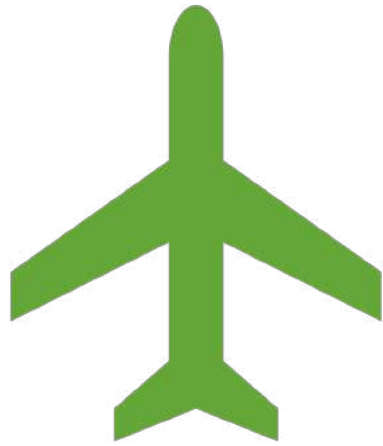
内閣官房情報通信技術（IT）総合戦略室長代理（副政府 CIO）
内閣官房番号制度推進室長（副政府 CIO）
内閣官房イノベーション推進室イノベーション総括官
内閣官房デジタル市場競争本部事務局次長
内閣府大臣官房審議官（科学技術政策・イノベーション担当）
内閣府知的財産戦略推進事務局局長
個人情報保護委員会事務局審議官
総務省大臣官房総括審議官（情報通信担当）
文部科学省大臣官房サイバーセキュリティ・政策立案総括審議官
経済産業省商務情報政策局長

目次

1. なぜ、データを使う必要があるのか
2. 公共データの共有がもたらす社会的インパクト
3. 政策立案のためのデータ分析の流れ
4. データ活用におけるよくある勘違い
5. おすすめの文献

1. なぜ、 データを使う 必要があるのか

どちらの事故の方が怖いと感じますか？

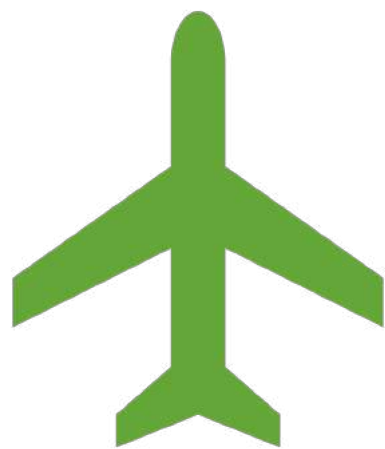


飛行機事故



自動車事故

実際に死亡事故に遭う確率を調べてみると...



約**252**万件に1回
起こる確率

飛行機事故

0.00004%

2018年に世界で発生した飛行機事故件数：15 / 年間総フライト数：37,800,000
出典元：Aviation Safety Network



約**3.5**万件に1回
起こる確率

自動車事故

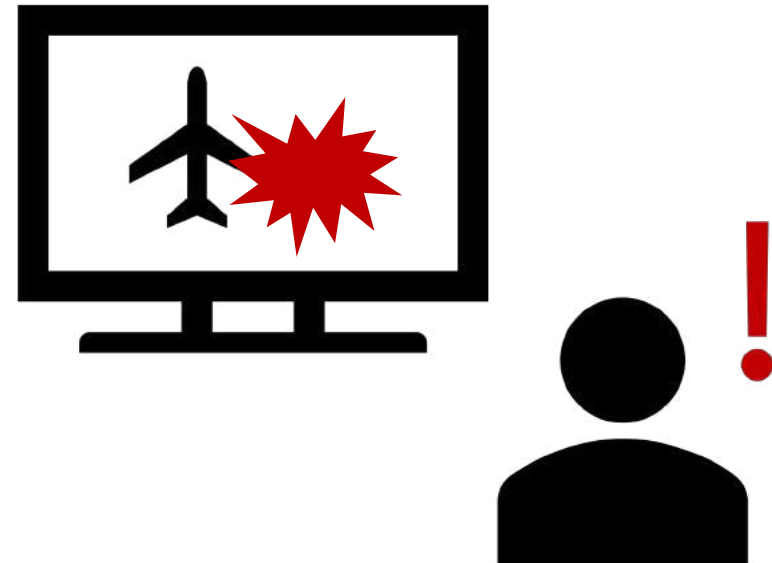
0.00279%

2018年の国内の自動車交通事故による死亡者数：3,532 / 総人口：126,435,000
出典元：警視庁, 総務省統計局

飛行機の
約**70倍**

「可用性ヒューリスティック」と呼ばれる判断の傾向が原因

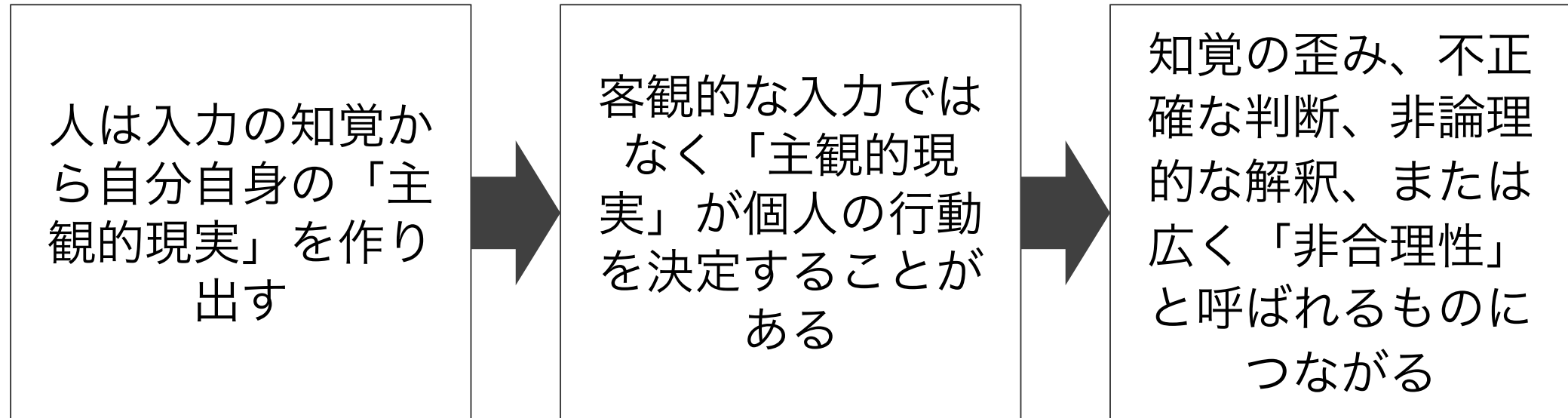
- 人は、ある事例を思い浮かべやすければ、起こりやすいと判断しやすい傾向がある
- 認識、理解、決定の際に、**思い出しやすい情報だけに基づいて判断する傾向を**
「可用性ヒューリスティック (Availability heuristic)」と呼ぶ
- 人が陥りやすい「認知バイアス」のうちの1つ



飛行機事故は大きく報道されることが多く、印象に残りやすい

認知バイアス =無意識な思考の偏り・誤り

認知バイアス (cognitive bias) とは、
人が規範や合理性から外れた判断をしてしまう時の典型的なパターン。



出典：https://en.wikipedia.org/wiki/Cognitive_bias より翻訳

認知バイアスの例①： アンカリング効果

○ アンカリング効果

人間は商品を相対的な比較でしか判断することができない

アンカリング効果なし



価格: ¥ 12,000

ん～高いなあ.....
買うのどうしようかな
他も検討するか...



アンカリング効果あり



参考価格: ~~¥ 56,000~~

価格: ¥ 12,000

お！普通より安く
なっている！
買おう！



Twitter:@never_be_a_pm
<https://dividable.net>

認知バイアスの例②： サンクコスト（埋没費用）

○ 埋没費用

回収できない投資額を回収しようとしてしまう認知バイアス

埋没費用なし

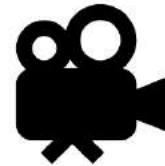


チケット価格: ¥0

この映画つまらない
から見るのやめよう



埋没費用あり



チケット価格: ¥1000

1000円払ってしまったし、もったいない
から見つけよう...



Twitter:@never_be_a_pm
<https://dividable.net>

どうすれば認知バイアスを外せるか？



どんな認知バイアスがあるか
知っておけば気を付けられる...？

名前の付いている認知バイアスは190以上

<http://lelang.sites-hosting.com/naklang/method.html>

認知バイアス一覧で社会心理学入門
～社会科学の知の蓄積を活用した社会教育の実現に向けて～

曉 美焰 (Xiao Meiyao) 社会学研究家, 2019.7.18 祝2.6版完成!

疑似科学を生み出すのは人間の思考が本来持っている誤りやすい傾向である。それ故に「科学と疑似科学の境界」を判断するためには、社会科学の知識は避けて通れない。ここでは人間の誤りやすい傾向について、人類の英知である「社会心理学」の偉大なる成果である認知バイアス一覧を英語版に基づいて紹介する。ここに紹介する知識は知っておくだけで人生に役立つ知識である事に間違いない。これらの概念を紹介する日本語のウェブサイトを探すのが難しい事自体が、日本人が論理的思考をしていない事の証明であろう。これまで社会科学とは縁の無かった科学技術系の人達が、少しでも社会科学に興味を持っていただく事を祈る。

1. 基本用語

- 帰属** 出来事や他人の行動や自分の行動の原因を説明する心的過程 (...のせいにする)。簡単にいえば、人柄のせいにするのが内的帰属であり、事情のせいにするのが外的帰属である。
- 動機づけ** 行動を始発させ、目標に向かって維持・調整する過程・機能。好奇心や関心によってもたらされ、賞罰に依存しない行動が内発的動機づけであり、義務、賞罰、強制などによってもたらされる行動が外発的動機づけである。

以下の認知バイアスの日本語訳は、できるだけ既に翻訳されている用語を用いている。しかし、翻訳が見つからなかった場合には勝手に日本語を造語し

2. 信念形成の際の認知バイアス一覧

- 後知恵バイアス (Hindsight bias)**
物事が起きてからそれが予測可能だったと考える傾向。
- 可用性カスケード (Availability cascade)**
主張を何度も聞いているうちに、真理であると確信する傾向。
- 可用性ヒューリスティック (Availability heuristic)**
認識、理解、決定の際に、思い出しやすい情報だけに基づいて判断する傾向。
- 観察者期待効果 (Observer-expectancy effect)**
観察者が期待する効果を観察する観点で解析し、無意識のうちにデータを誤って解釈する傾向。観察者効果の一種。
- 感情移入ギャップ (Empathy gap)**
怒ったり恋愛したりしている時に、その感情を持たない視点で考える事ができない傾向。

常に全てを
意識するのは
無理!!

認知バイアスは避けられない

自分も、他人も、認識や判断には

常に認知バイアスが
かかっている可能性がある

という前提でいる必要がある



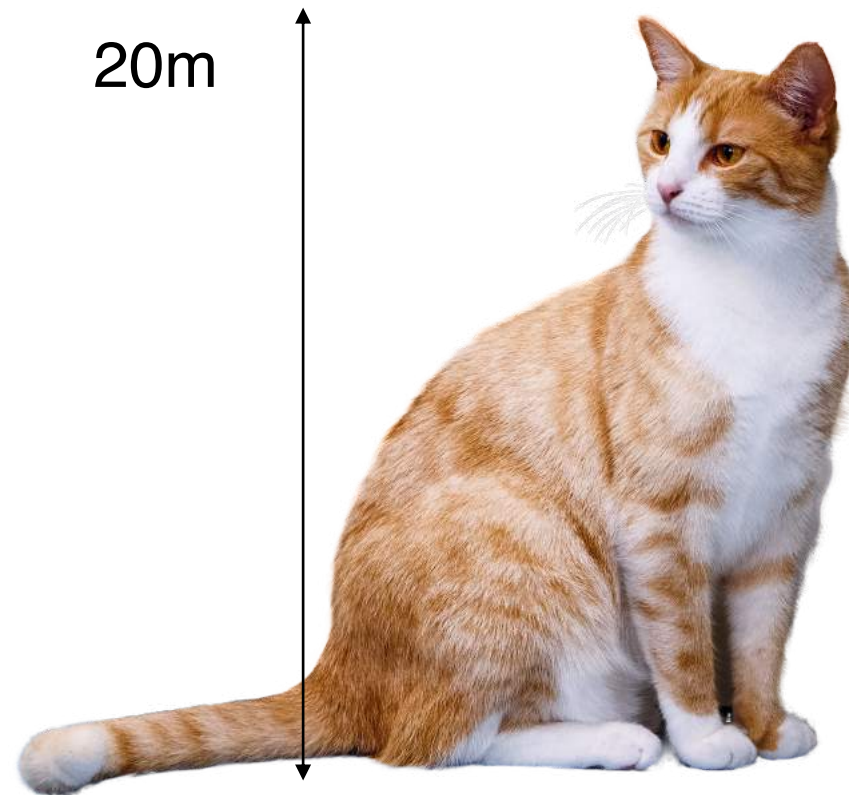
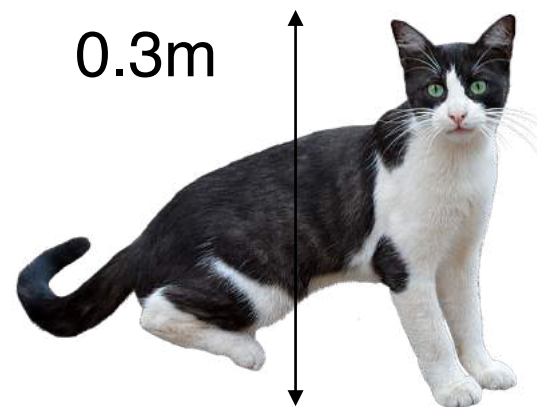
認知バイアス対策

どうすればバイアスがかかっていることに気付き、
正しい判断に近付くことが出来るのか？



主観的な情報だけでなく、客観的な情報を使う

客観的 = 誰が見ても大体同じ解釈ができる



客観的

Q1.
どちらの猫が大きい？

主観的

Q2.
どちらの猫が可愛い？

数値を使うと解釈の齟齬を減らせる

例) ある市における救急隊員が抱えていた課題

配属先が変わって
業務負担が増えた



しかし上司の理解が得られない



数値を使うと解釈の齟齬を減らせる

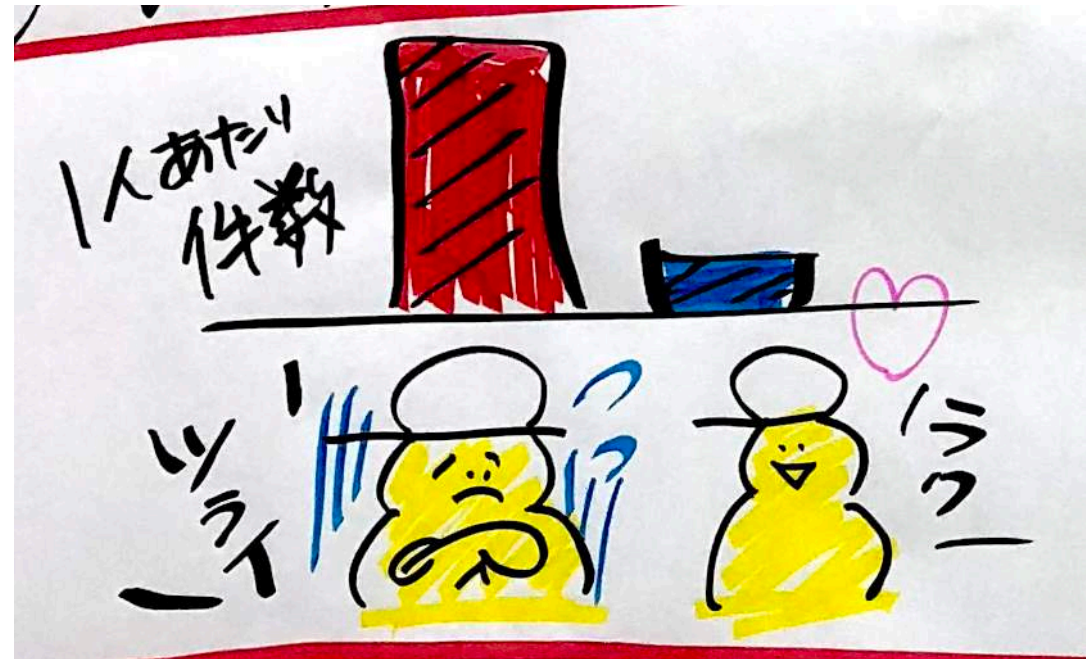
市内の各隊の隊員1人あたりの
出動件数*を算出して比較

※使用データ：消防年報の年間救急出動件数÷各隊の配属人数



最大で **7.75倍** の差

- 数字は言葉よりも解釈の幅が狭い
- 解釈の齟齬が言葉に比べて起こりにくい



出典元：Code for Japan データアカデミーにおける某自治体のチームの当日発表資料

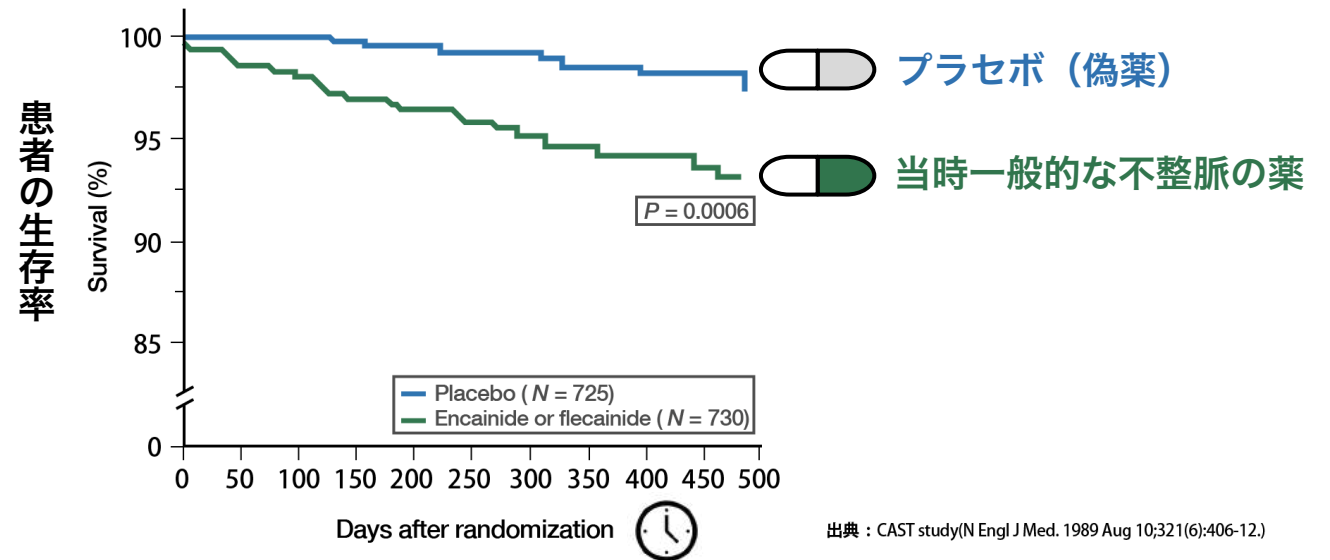
経験やカンだけではなく、それを裏付ける客観的なデータが必要

■ EBPM : Evidence Based Policy Making

...エビデンス (科学的根拠) に基づいた政策立案

■ EBPMは医学 (Evidence Based Medicine) から派生した考え方

- 1989年、当時一般的に使用されていた不整脈の薬の効果をデータを取って検証したところ、服用によって死亡率が高まることが判明



経験的知識のみで判断することの危険性

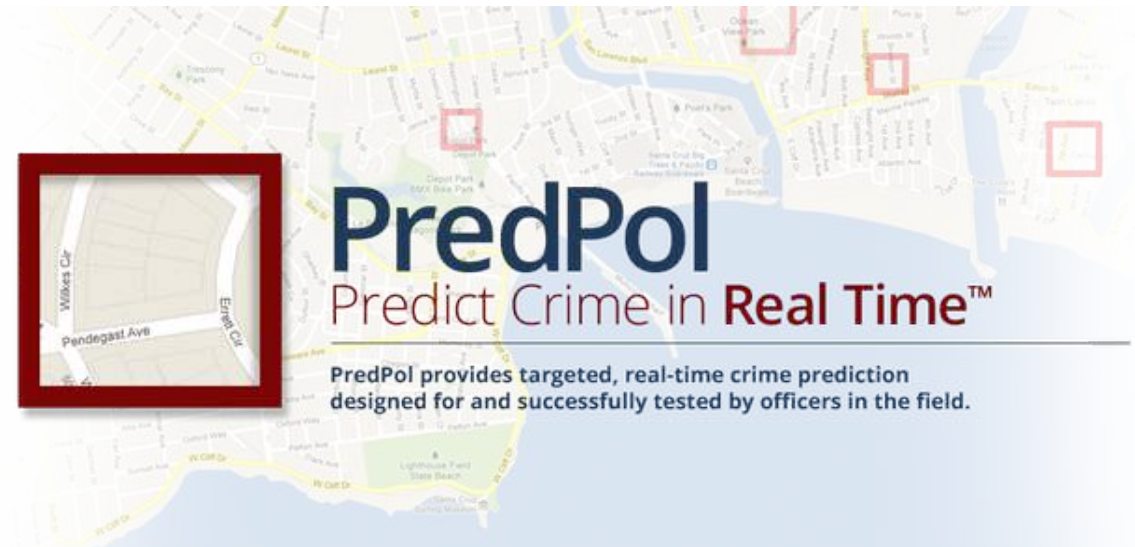


データに 基づかない施策は 民間療法と同じ

データを活用せずに、
経験や勘や慣例のみで施策を進めるのは、
民間療法で病気を治そうとするようなもの

データ分析結果が経験やカンを越えた例： PredPol（米国の犯罪発生予測システム）

- PredPol はカリフォルニア大学の研究グループが開発した犯罪予測システム
- 過去の膨大な犯罪情報のデータベースをもとに、将来「いつ、どこで」犯罪が起きるのか、機械学習をベースに予測
- ロサンゼルス市警のプロの犯罪分析官とPredPolの予測を比較したところ、犯罪分析官に比べ**2倍以上**PredPolの方が正確に予測できた



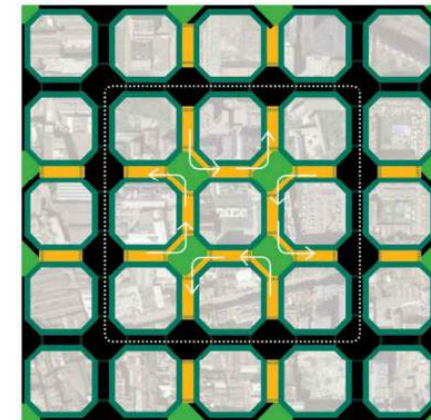
<https://www.predpol.com/>

従来の経験やカン頼りの方法ではなく、
データに基づく客観的な根拠から判断する方法へシフト

データで根拠を示すことで合意形成を進めた例： スーパーブロック構想（バルセロナ市）

- バルセロナでは自動車の交通量が多く、汚染物質やCO2、騒音などによる公害が課題
- 「市民中心都市」をコンセプトにかかげ、段階的に車が通れる道を制限する計画を実施
- 交通量や環境に関するセンシングデータを用いて分析を行い、計画が進んだ場合にどのくらい問題が改善するのか、具体的な数値で示すことで**合意形成を推進**
 - 交通量が**21%**減
 - **94%**の市民が危険なレベルの粒子状物質に晒されることはなくなる
 - **73.5%**の市民は「65デシベル以上の騒音」を経験せずにすむ

<https://citiesofthefuture.eu/superblocks-barcelona-answer-to-car-centric-city/>



Phase 1



Phase 2

2. 公共データの 共有がもたらす 社会的インパクト

サンフランシスコ市 × Yelp から見える データがもたらす社会変革



- 世界最大級のローカルビジネスのレビューサイト
- 日本でいうところの「食べログ」的なサービス

- サンフランシスコ市では、飲食店に対して実施した保健衛生検査の結果をデータとして公開している
- Yelpはそれをサービスに取り込み、“Health Score”として100点満点のスコア化して各飲食店のレビューページに表示



サンフランシスコ市 × Yelp から見える データがもたらす社会変革

Fog Harbor Fish House Claimed

★★★★☆ 4550 reviews

☆☆ Seafood, Bars

Pier 39
Ste A-202
San Francisco, CA 94133
Fisherman's Wharf, North Beach/Telegraph Hill

Watch video

Make a Reservation

Health Score 87 out of 100

衛生スコア：
87点

サンフランシスコ市 × Yelp から見える データがもたらす社会変革

Fog Harbor Fish House

November 20, 2017 – Routine Inspection

Violations

- Improper storage use or identification of toxic substances
- High risk food holding temperature [date violation corrected: 12/4/2017]
- Improper food storage
- Wiping cloths not clean or properly stored or inadequate sanitizer [date violation corrected: 11/20/2017]

Inspections

Date	Inspection Type	Violations	Score
November 20, 2017	Routine	4	87
July 17, 2017	Routine	2	86
April 25, 2016	Routine	3	92

直近の保健衛生
検査において違反
があった項目

Health Score

87

out of 100

About Health Scores

We collect public inspection data directly from your local health department. Due to the local health department's inspection schedule as well as the time it takes to pass that information on to us, it is possible that we may not display the most recent inspection data.

Please report data inaccuracies via one of the methods below:

- [Email](#)
- [Website](#)

Yelpはユーザーにより安心・安全を届けられる
サービスを提供可能に

サンフランシスコ市 × Yelp から見える データがもたらす社会変革



もしサンフランシスコ市が
衛生調査結果をサイトで公表している
だけだったら？

一部の人しか見に来ない

強み：市民の安全を守るために実施
した調査結果として、公平なデータ
を提供できる



もしYelpが
独自に飲食店を評価する
スコアを付けていたら？

公平性が担保されない

強み：店舗の営業データや口コミ
データを集めて、多くのユーザが
集まる場を持っている

データが共有されることで
「組織を超えて得意分野を分担できる社会」になる

行政のデータは誰にでも使えるように公開されている必要がある



もしサンフランシスコ市が
Yelpだけに衛生調査結果データを
渡していたら？



不公平に対し非難が集中

- 衛生調査は市の税金を使って運用されているので、納税者は等しくそのデータを使う権利を持っている
- 誰でも自由に使えるデータである必要がある

オープンデータ

オープンデータの定義

国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工、編集、再配布等）できるよう、次のいずれの項目にも該当する形で公開されたデータをオープンデータと定義する

【出典】「オープンデータ基本指針」平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定

1. 営利目的、非営利目的を問わず二次利用可能なルールが適用されたもの
2. 機械判読に適したもの
3. 無償で利用できるもの

“Open By Default”

2013年 G8 「オープンデータ憲章」 による国際的な合意



「税金を使って作られた
データは全て公共財として
公開するべきである」
という考え方を
原則として採用

“オープンデータ”とは、
元々皆さんのものだったデータを
返すことだと考えています。

- Miquel Mateu, Open Data Manager of Barcelona City



14-18 June 2016, World Data Viz Challenge 2016,
Barcelona

2016年：官民データ活用推進基本法施行

→国及び自治体はオープンデータに取り組むことが義務付けられた

官民データ活用推進基本法

目次

- 第一章 総則（第一条—第七条）
 - 第二章 官民データ活用推進基本計画等（第八条・第九条）
 - 第三章 基本的施策（第十条—第十九条）
 - 第四章 官民データ活用推進戦略会議（第二十条—第二十八条）
- 附則

■第一章 総則

（目的）

第一条 この法律は、インターネットその他の高度情報通信ネットワークを通じて流通する多様かつ大量の情報を適正かつ効果的に活用することにより、急速な少子高齢化の進展への対応等の我が国が直面する課題の解決に資する環境をより一層整備することが重要であることに鑑み、官民データの適正かつ効果的な活用（以下「官民データ活用」という。）の推進に関し、基本理念を定め、国、地方公共団体及び事業者の責務を明らかにし、並びに官民データ活用推進基本計画の策定その他官民データ活用の推進に関する施策の基本となる事項を定めるとともに、官民データ活用推進戦略会議を設置することにより、官民データ活用の推進に関する施策を総合的かつ効果的に推進し、もって国民が安全で安心して暮らせる社会及び快適な生活環境の実現に寄与することを目的とする。

出典：https://www.kantei.go.jp/jp/singi/it2/hourei/detakatsuyo_honbun.html

（条文より抜粋）

（国及び地方公共団体等が保有する官民データの**容易な利用等**）

第十一条 国及び地方公共団体は、自らが保有する官民データについて、個人及び法人の権利利益、国の安全等が害されることのないようにしつつ、国民がインターネットその他の高度情報通信ネットワークを通じて容易に利用できるよう、必要な措置を講ずるものとする。

（情報システムに係る規格の整備及び**互換性の確保等**）

第十五条 国及び地方公共団体は、官民データ活用に資するため、相互に連携して、自らの情報システムに係る規格の整備及び互換性の確保、業務の見直しその他の必要な措置を講ずるものとする。

オープンデータの意義

1. 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
2. 行政の高度化・効率化
3. 透明性・信頼の向上

【出典】 「オープンデータ基本指針」 平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定

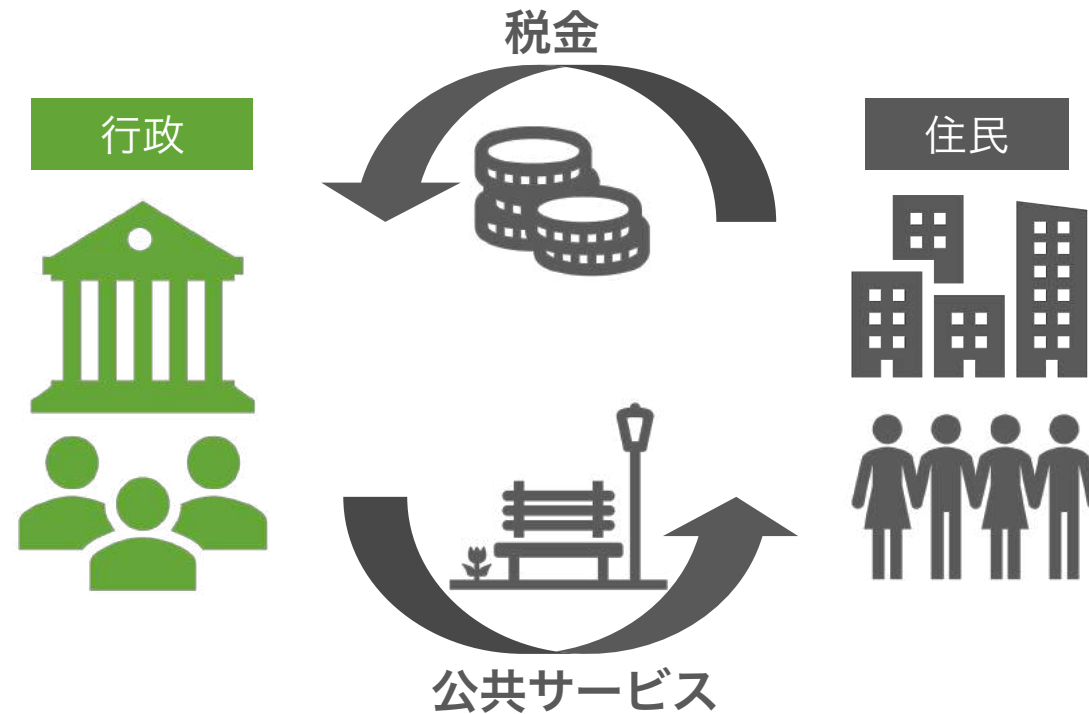
オープンデータの意義

1. 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
2. 行政の高度化・効率化
3. 透明性・信頼の向上

【出典】 「オープンデータ基本指針」 平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定

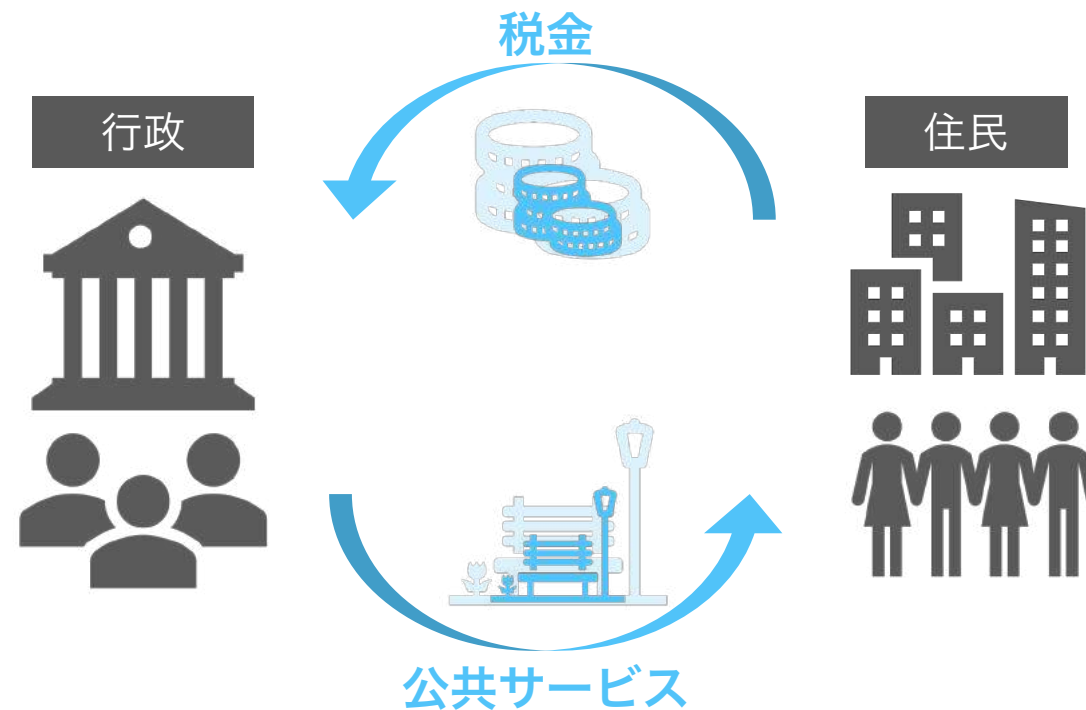
なぜ官民協働が必要なのか？

これまでは行政が公共サービス運営の主体だった

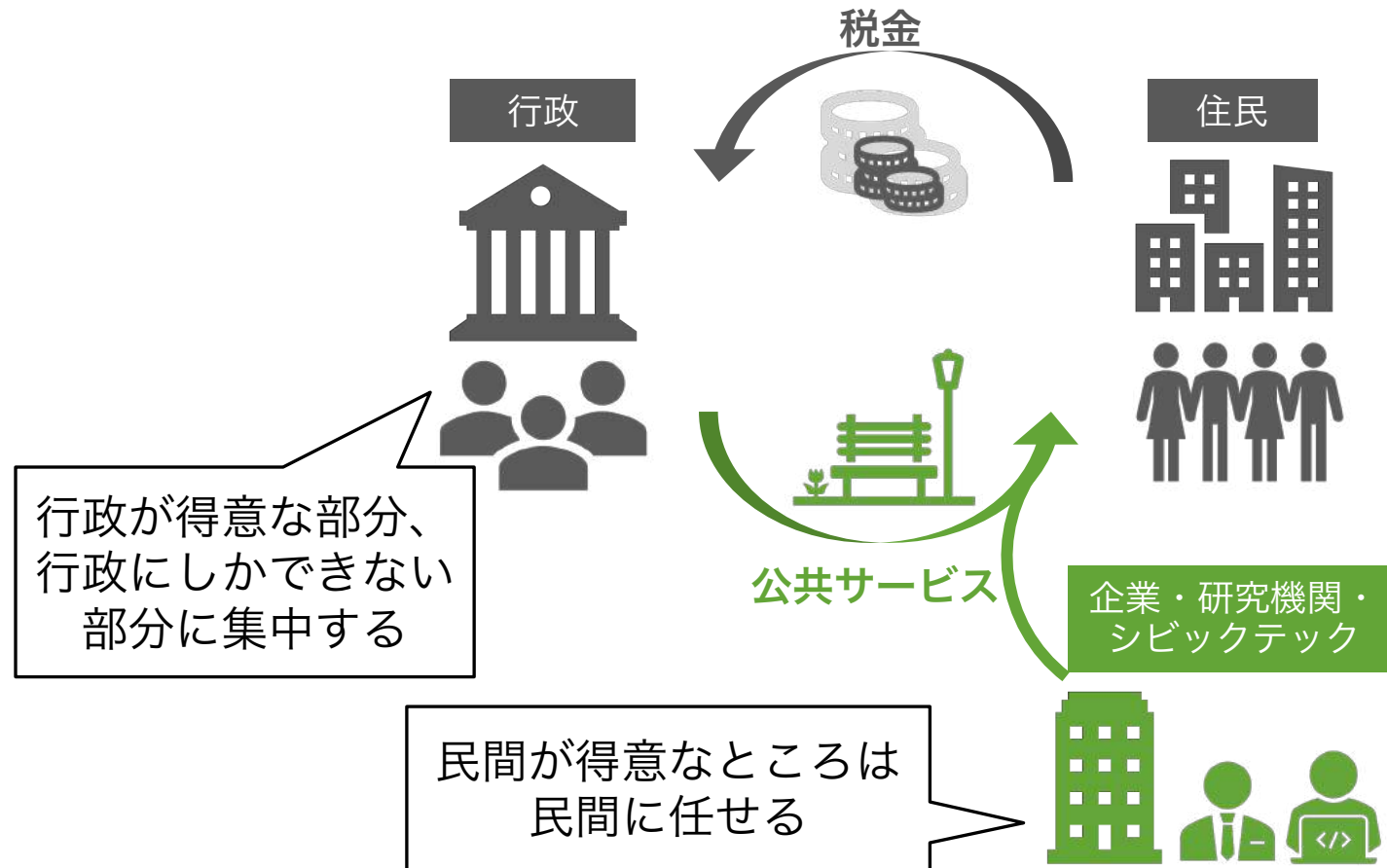


なぜ官民協働が必要なのか？

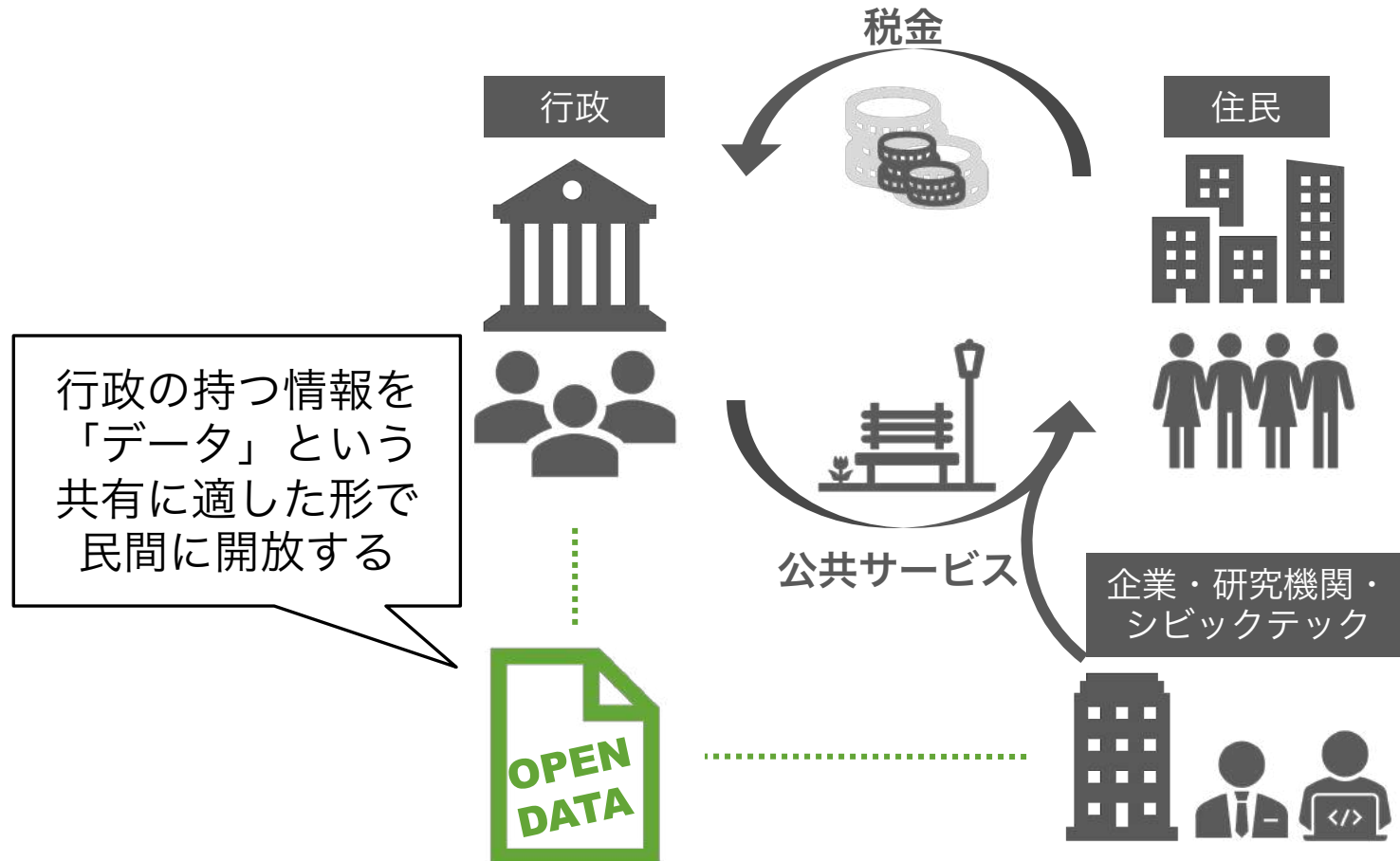
税収が減ると、これまで通りの公共サービスは維持できなくなる



官民協働によって地域を持続可能にする



オープンデータによって、 組織を超えて得意分野を分担できる社会 が実現する



官民協働の事例

東京都新型コロナウイルス感染症対策サイト

- 東京都がオープンソースとして公開後、全都道府県に展開
- データをわかりやすく可視化して見せるデザインに定評
- 約300人のエンジニアが開発に協力

<https://stopcovid19.metro.tokyo.lg.jp/>

東京都
新型コロナウイルス感染症対策サイト

多言語対応選択メニュー
Lang: 日本語

都内の最新感染動向

新型コロナウイルス感染症が心配なときに

新型コロナウイルスの感染が判明した方へ

ご家庭でのマスク等の捨て方

お子様をお持ちの皆様へ

都民の皆様へ

企業の皆様・はたらく皆様へ

東京都新型コロナウイルス感染症対策本部報

東京都 新型コロナウイルス

都内の最新感染動向 最終更新 2020/05/14 19:15

最新のお知らせ

東京都緊急事態措置について

- 2020-05-14 [【最新情報動画】知事から都民の皆様へ\(5/14\)](#)
- 2020-05-14 [等決処分によるマスク及び個人防護具（ガウン等セット）等の買入れについて](#)
- 2020-05-14 [新型コロナウイルス感染症患者公表数の修正について](#)

自分や家族の症状に不安や心配があればまずは電話相談をどうぞ

[相談の手順を見る](#)

検査陽性者の状況

- (注) チャーター機帰国者、クルーズ船乗客等は含まれていない
- (注) 「重症」は、集中治療室（ICU）等での管理又は人工呼吸器管理が必要な患者数を計上
- (注) 「退院等」には、療養期間経過を含む。
- (注) 退院者数の把握には一定の期間を要しており、確認次第数値を更新している



2020/05/14 19:15 更新

検査実施状況

- (注) 検査実施人数には、医療機関等による検査人数、チャーター機帰国者、クルーズ船乗客等は含まれていない。
- (注) 速報値として公開するものであり、後日確定データとして修正される場合あり



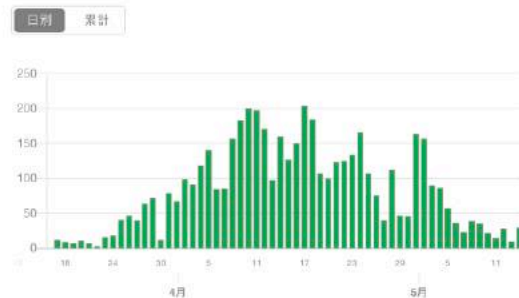
2020/05/14 18:45 更新

新規患者に関する報告件数の推移

30人

5/14日別値 (前日比: +20人)

- (注) 保健所から発生届が提出された日を基準とする
- (注) 医療機関等が行った検査も含む
- (注) チャーター機帰国者、クルーズ船乗客等は含まれていない



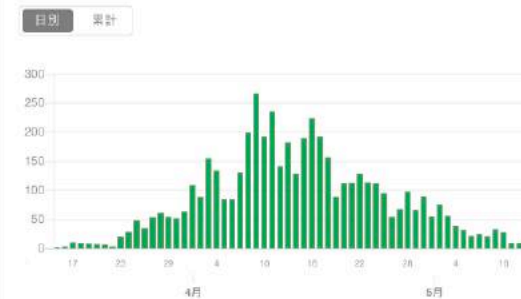
> テーブルを表示

オープンデータを手入
2020/05/14 19:15 更新

PCR検査陽性者の発生動向 (確定日別による陽性者数の推移)

17人

5/13日別値 (前日比: +7人)

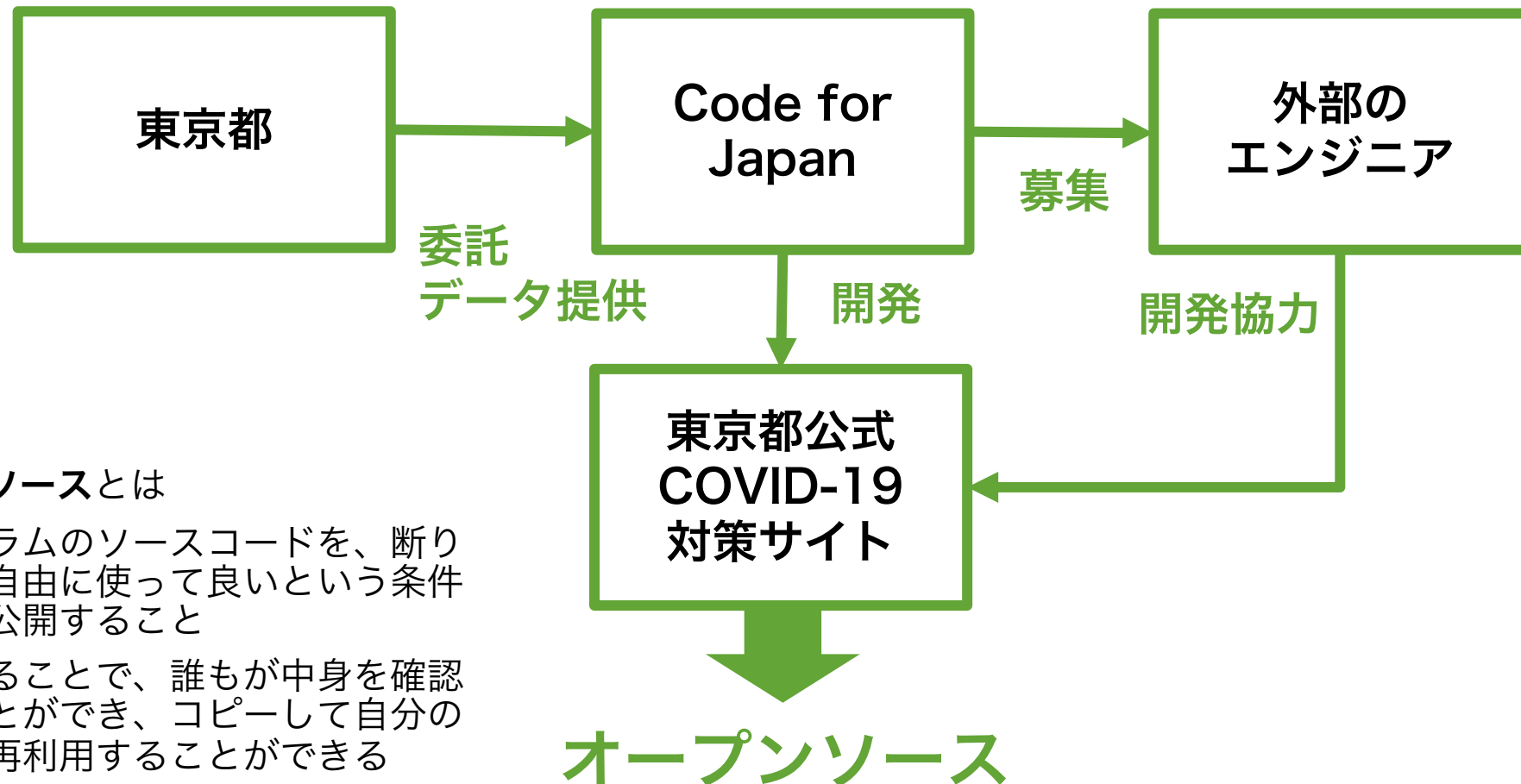


- (注) 患者発生の動向をより正確に分析するため、各保健所から報告があった患者の発生情報を、PCR検査により陽性であることを医師が確認した日別（確定日別）に整理したものである。

> テーブルを表示

2020/05/14 19:15 更新

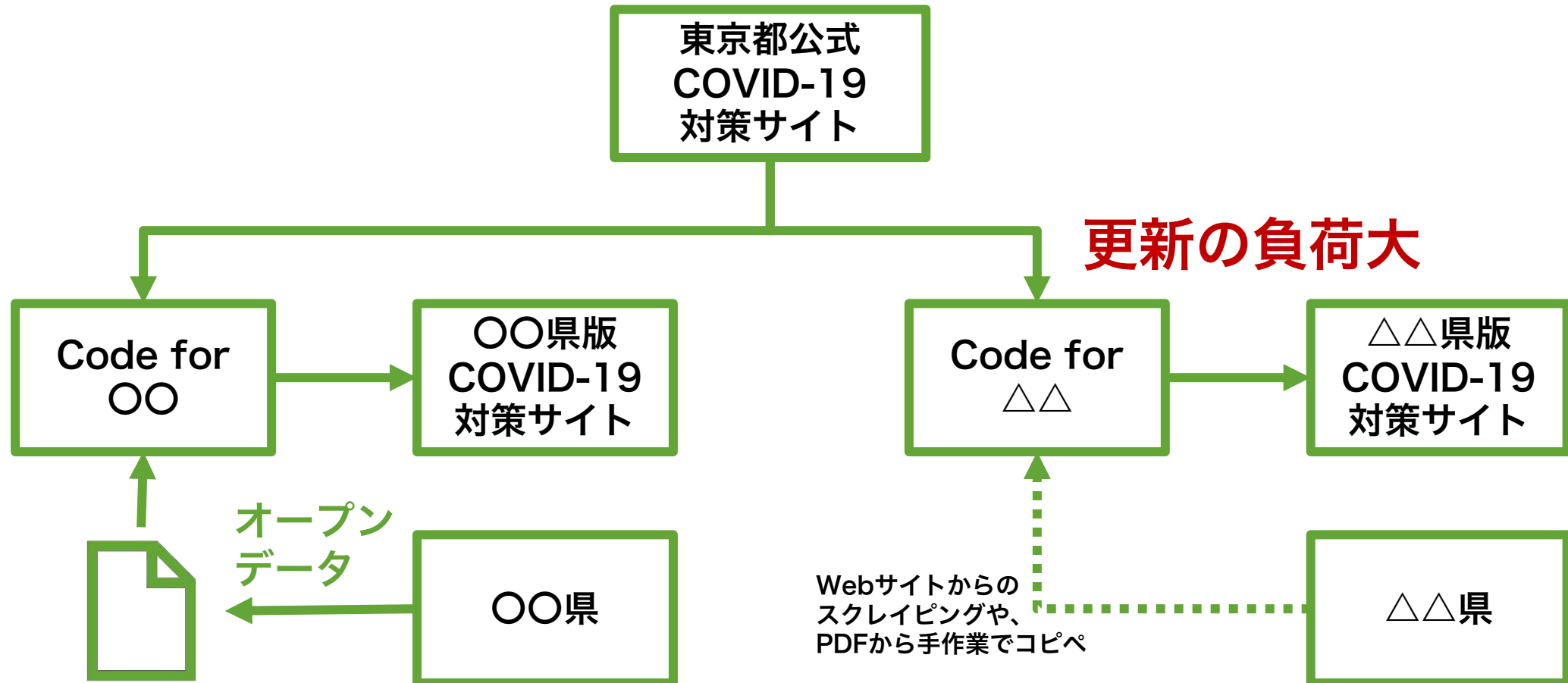
3月上旬、COVID-19対策サイトを 開発しオープンソース*化



*オープンソースとは

- プログラムのソースコードを、断りなしに自由に使って良いという条件で広く公開すること
- 公開することで、誰もが中身を確認することができ、コピーして自分の目的に再利用することができる

シビックテック等により各都道府県に展開したが、 毎日のデータ更新に課題



新型コロナウイルス対策データの標準フォーマット公開

	A	B	C	D	E
1	新型コロナウイルス感染症対策に関するオープンデータ項目定義書				
2	・本データ項目定義書は、新型コロナウイルス感染症対策の一環として、各自治体において標準フォーマットでのオープンデータを公開できるように				
3	・どのような情報を優先的にオープンデータ化し、情報発信の効率化を図るべきか、指針を検討するために本書をご利用ください				
4	・利用者にとっての使いやすさを担保するためには標準化された形でのデータ公開が望ましいですが、必ずしも本書通りに作成する必要はなく、適				
5	・迅速かつ正確な情報発信を行うことを最優先してください				
6					
7	データセット一覧				
8	No	データセット名	説明	備考	
9	01	陽性患者属性	陽性患者の年代・性別・居住地等の属性情報を一覧化したデータ		
10	02	検査実施人数	PCR検査を受けた人の数を日別に集計したデータ		
11	03	検査実施件数	PCR検査の実施数を日別に集計したデータ		
12	04	陰性確認数	陰性確認が完了した件数を日別に集計したデータ	陰性確認とは、検査で陽性となった患者等について、症状改善後にウイルスの保有状態が継続しているかどうかを確認するための検査。検査陽性となり入院した患者は、一定間隔を置いて2回連続で陰性となった場合に退院が可能となる。	
13	05	コールセンター相談件数	コールセンターや受診相談窓口における相談件数を日別に集計した	コールセンター・受診相談窓口ともに同じフォーマットを利用可能	
14	06	支援制度	補助金や助成金、融資などの支援制度の情報を一覧化したデータ (※新型コロナウイルス感染症対策に限らず使用可能)	内閣官房情報通信技術（IT）総合戦略室、総務省及び経済産業省によって整備されたフォーマット（ https://www.meti.go.jp/covid-19/info-service.html ）に原則準拠しているが、国の支援制度をシステムから入力することを前提とした設計になっていたため、地方自治体でのデータ整備のしやすさやオープンデータとしての使い勝手を考慮し、一部仕様変更を行った。	

- 01 陽性患者属性
- 02 検査実施人数
- 03 検査実施件数
- 04 陰性確認数
- 05 コールセンター相談件数
- 06 支援制度

プロジェクトページ: <https://www.code4japan.org/activity/stopcovid19#>

総務省、内閣官房IT室、地域情報化アドバイザー、オープンデータ伝道師と協力して自治体に展開

各都道府県行政情報化担当課
各指定都市行政情報化担当課

御中

事務連絡
令和2年3月31日



総務省情報流通行政局
地域通信振興課

新型コロナウイルス感染症対策サイトのためのデータ公開について
(情報提供)

新型コロナウイルス感染症に対しては、政府として、感染症の発生状況や様々な支援制度についての情報提供を行っているところですが、各地方公共団体や各地のエンジニア有志などにより、地域ごとの情報提供サイトも立ち上げられています。(⇒「別添」P. 2～3参照)

オープンソースを活用した新型コロナ対策サイトの各地での公開

- 東京都が新型コロナウイルス対策サイトを開発し、オープンソースで公開（GitHub開発者向けプラットフォーム）。その後、東京都の当該サイトを起点として、各地のエンジニアが改善を進めている。
- オープンソースとして公開することで、その後、全国のエンジニア有志などが相次いで自分の地域のサイトを公開。（2020年3月26日現在、34都道府県、1地域（九州）、2政令市（千葉市、福岡市）で公開中。）

公開後、3週間の間に
224名が改善に協力し、
750件の提案がなされ、
671件が取り入れられた。

出所：東京都 新型コロナウイルス感染症対策サイト（閲覧日：2020-03-24）
<https://stopcovid19.metro.tokyo.lg.jp/>

出所：Code for Japan 開出の迅速性をとくに考慮。
(調査は開出時に、in.GitHub-1)

Q&A

Q 自治体の公式サイトを東京都の公式サイト等のオープンソースで作っても大丈夫ですか。

もちろん大丈夫です。今回の場合、東京都が公式サイトとして作ったものを、オープンソースで公開しており、いくつかの自治体では、これを使って公式サイトを開発しています。オープンソースは、随時改善されることが多いので、改善点を自分のサイトにも反映していくことが望ましいです。他方、一般公開されているオープンソースを活用される際には出元ライセンスにご留意ください。

Q 地元の有識者が作った非公式サイトで自治体のデータを公開しても大丈夫ですか。

自治体などがオープンデータで公開している情報は、誰でも自由に使うことができます。オープンデータでなくても、今回のように多くの人に知ってもらいたい情報は、正しく伝わるのであれば、より多くの手段で住民の方々などに届けることが望ましいです。統計データなどは、基本的には著作権が発生しないので、誰でも自由に使うことが出来ます。オープンデータと明示すれば、使用料はより安心して使うことができます。データを掲載する際に出所を明記し、リンクを張って元データにたどり着けるようにしましょう。

Q シドックテックが作成した「新型コロナウイルス感染症対策に関するオープンデータ項目定義書」を使ってデータを作ったのですが、どのように公開すればよいですか。

まずは自分の自治体のオープンデータサイトで公開してください。
今後、内閣官房IT総合戦略室が、各自治体の公開状況を把握して、まとめて公開する予定です（時期は未定）。「新型コロナウイルス感染症対策に関するオープンデータ項目定義書」に準拠してデータを作成・公開した場合、内閣官房IT総合戦略室に連絡すれば、同室のサイトでもリンクが貼られます。

Q オープンデータやサイト構築に関する相談をしたいときは、どこに連絡すればよいですか。

下記の相談窓口にお気軽にご相談ください。

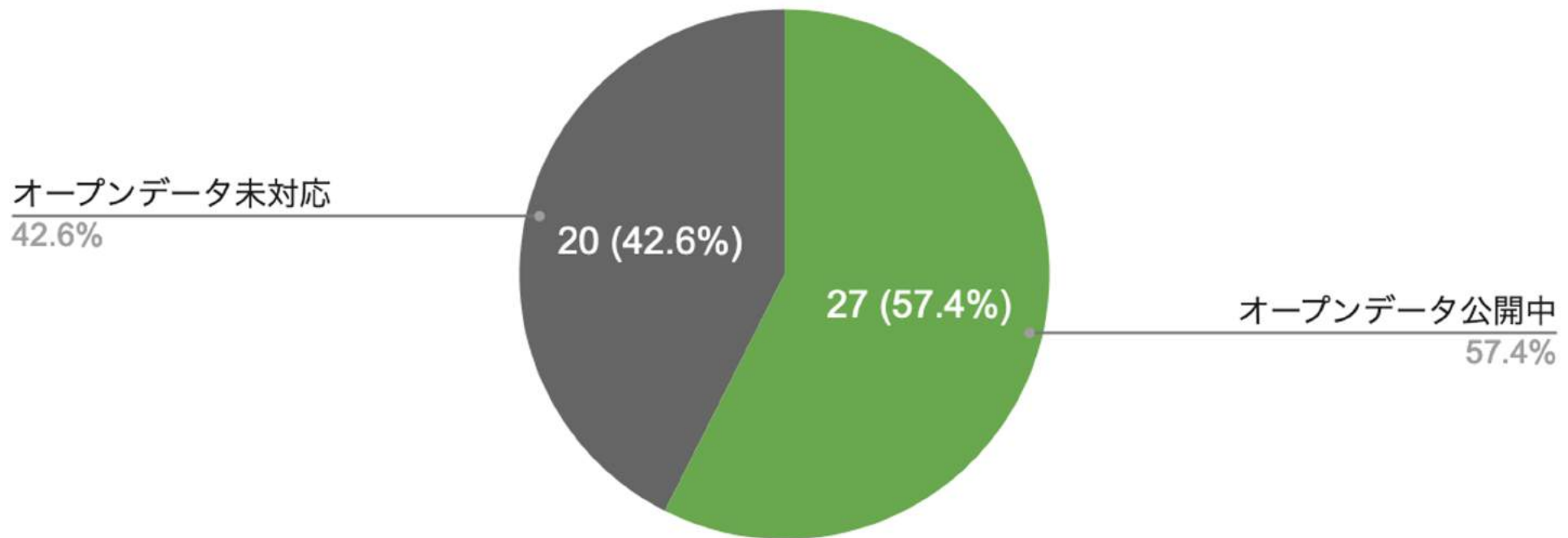
総務省：オープンデータ相談サイト
<https://www.opendata-howto.org/>

総務省：官民データ、統計データに関する相談窓口
https://www.soumu.go.jp/menu_sesaku/gyoumukanri_sonota/kanmin_data/index.html

内閣官房IT総合戦略室：官民データに関する総合的な相談窓口
<https://cio.go.jp/it-sodan/>

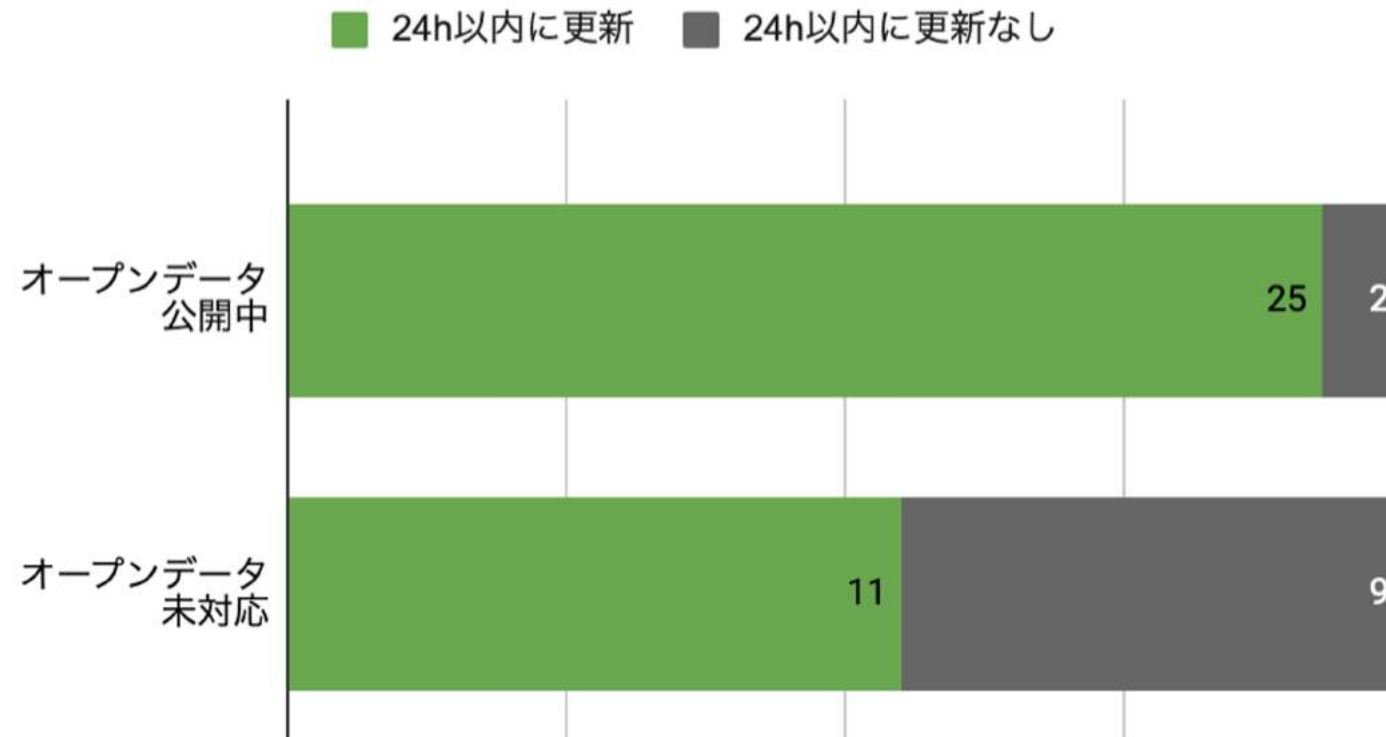
2020年7月21日時点、 オープンデータ化対応都道府県は半数超

COVID-19感染者関連データをオープンデータ化している都道府県



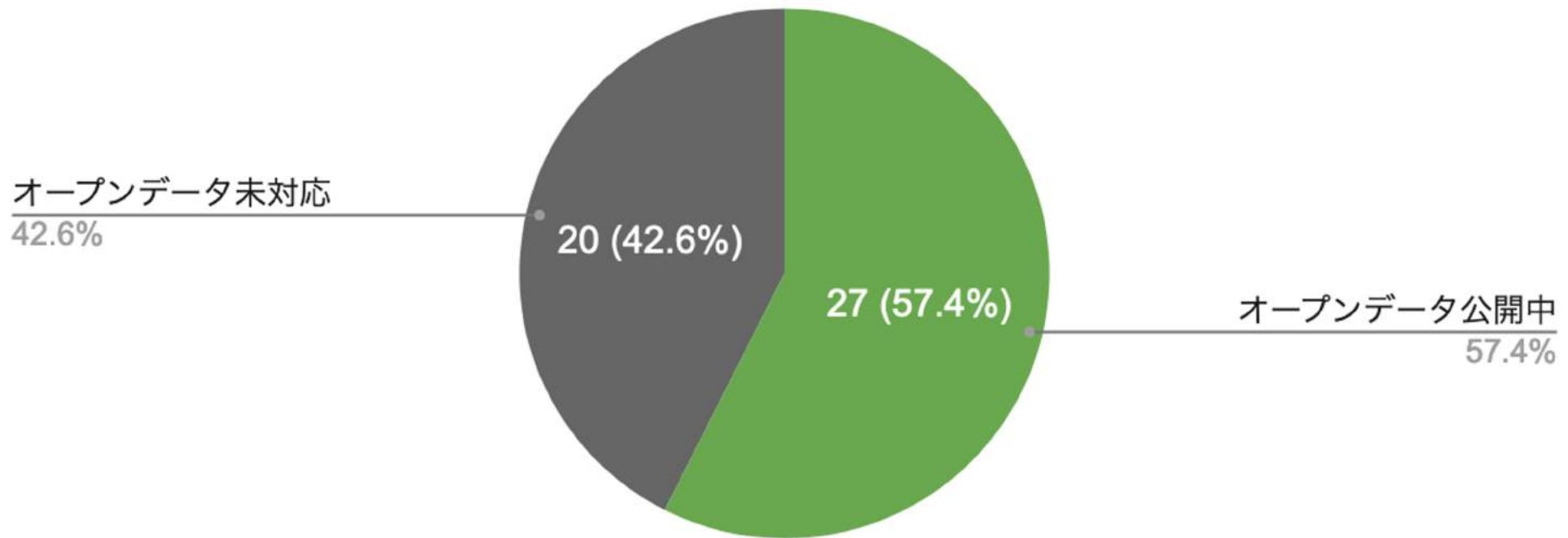
オープンデータの有無によって サイト更新速度に違いが生じていた (2020年7月21日時点)

新型コロナウイルス感染症対策サイトで24h以内に更新のあった都道府県の割合の比較



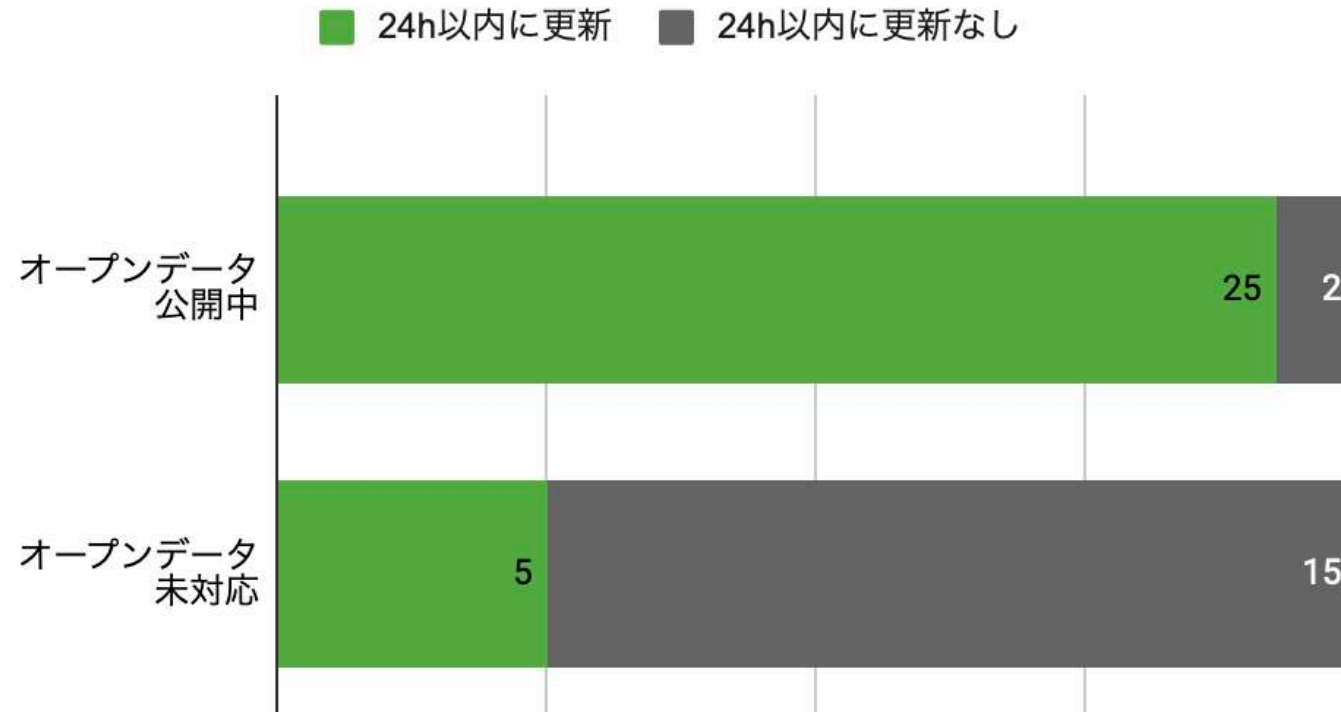
2020年11月8日時点、 オープンデータ化対応都道府県は変わらず

COVID-19感染者関連データをオープンデータ化している都道府県



オープンデータの有無によって サイト更新速度の差が広がった (2020年11月8日時点)

新型コロナウイルス感染症対策サイトで24h以内に更新のあった都道府県の割合の比較



更新停止状態のままのサイトや、 閉鎖するサイトも



栃木県

栃木県
新型コロナウイルス感染症
対策サイト

栃木県内の最新感染動向

新型コロナウイルス感染症が
心配なときに

栃木県新型コロナウイルス感
染症に関する情報

栃木県主催等 中止又は延期
するイベント等

知事からのメッセージ

当サイトについて

栃木県公式ホームページ

Lang: 日本語



このサイトの内訳物は [クリエイティブ・コモンズ 表示 4.0 ライセンス](#) の下に提供されています。

栃木県内の最新感染動向 最終更新 2020/10/31 09:59

閉鎖のお知らせ

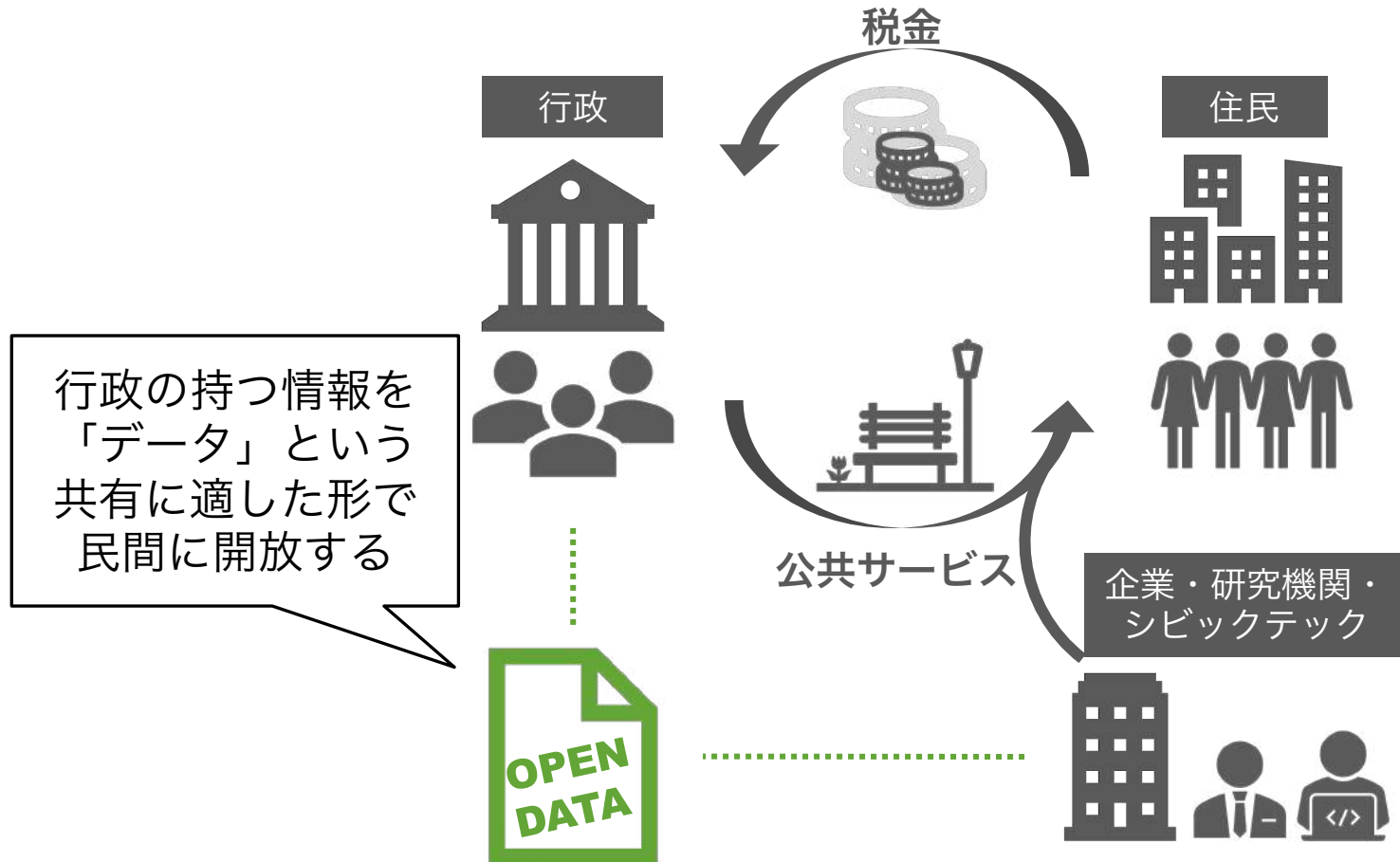
当サイトは2020年10月末をもって閉鎖いたしました。

当初は新型コロナ感染症に関して見やすくまとまった情報が少なく、当サイトはそこをサポートすべく有志で立ち上げられたものでした。しかし、現在、[栃木県の公式サイト](#)でも見やすくまとまった情報が提供されるようになってきました。そのため、本サイトの一定の役割は終えたものと判断し閉鎖することとなりました。

今まで当サイトをご利用頂きましてありがとうございました。

今後は[栃木県の公式サイト](#)で情報を確認頂きますよう、よろしくお願いいたします。

組織を超えて得意分野を分担できる社会の実現のために、国や自治体の保有するデータの公開は急務



オープンデータの意義

1. 国民参加・官民協働の推進を通じた
諸課題の解決、経済活性化

2. 行政の高度化・**効率化**

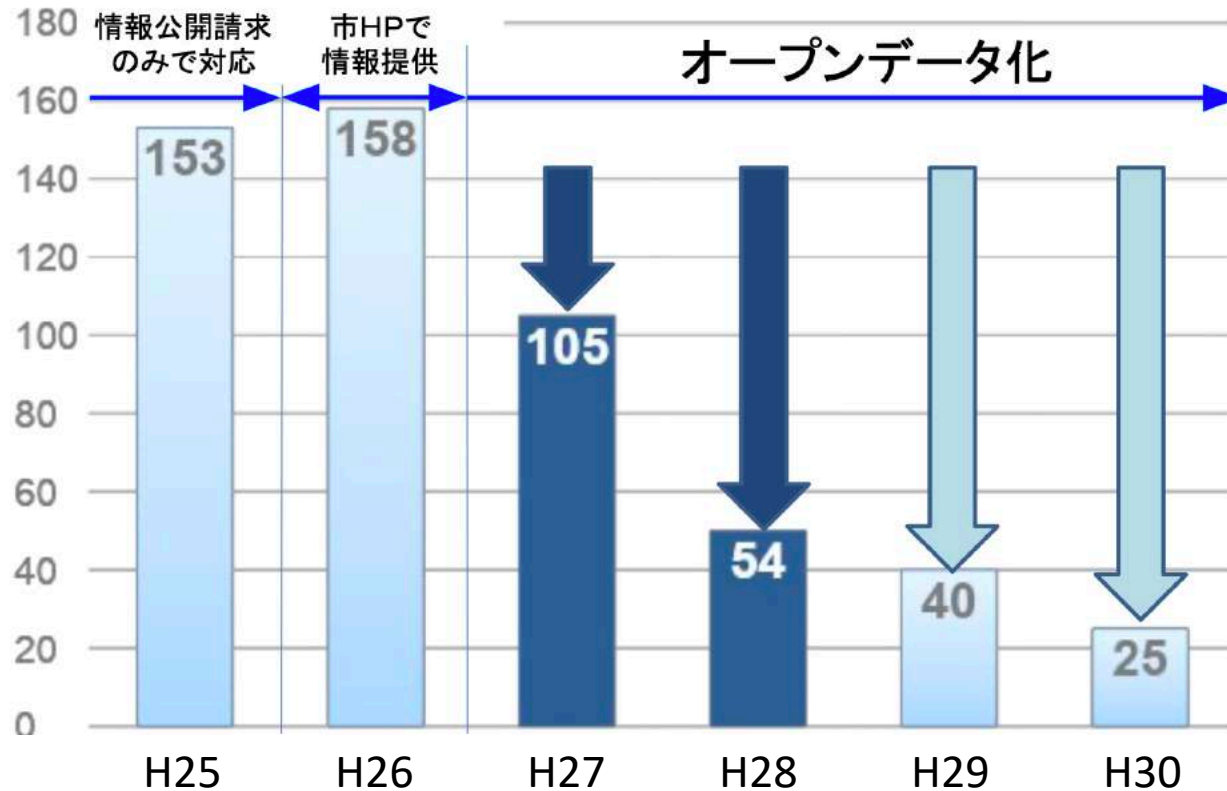
3. 透明性・信頼の向上

オープンデータ化
する作業の分、
仕事が増えるのでは…？



【出典】 「オープンデータ基本指針」平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定

事例： 情報公開請求の軽減（静岡市）



- 情報公開請求件数の多かった食品衛生許可に関するデータをオープンデータとして公開
- 情報公開請求の件数が減り、業務負担の軽減に成功

年間 **100件** 減
↓
約 **300時間**
業務時間削減

出典：新庄 大輔氏（静岡市総務局ICT推進課）「知的資産経営WEEK_OpendataAPI」
<https://www.slideshare.net/DAISUKESHINJO/weekopendataapi-227515360/1>

オープンデータの意義

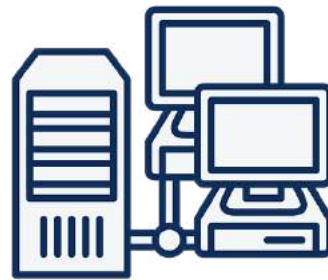
1. 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
2. 行政の高度化・効率化
3. 透明性・信頼の向上

【出典】 「オープンデータ基本指針」 平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定



透明性・信頼の向上は “共創”を進めるために必要

- 共創 = Co-creation
- 共創を進めるためには、行政と民間が対等な立場で話ができる必要がある
- そのためには、行政と民間の「情報の非対称性」を解消する必要がある
- 客観的なデータを共有して、同じものを見ながら話ができるのがベスト



オープンデータは
庁内でデータを適切に管理できて
いるかを示す役割もある

- そもそも庁内でデータをきちんと管理できていなかったら、公開はできない
- 不正の抑止力にもなることが期待できる

オープンデータの意義

1. 国民参加・官民協働の推進を通じた諸課題の解決、経済活性化
2. 行政の高度化・効率化
3. 透明性・信頼の向上

【出典】 「オープンデータ基本指針」 平成29年5月30日
高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議決定

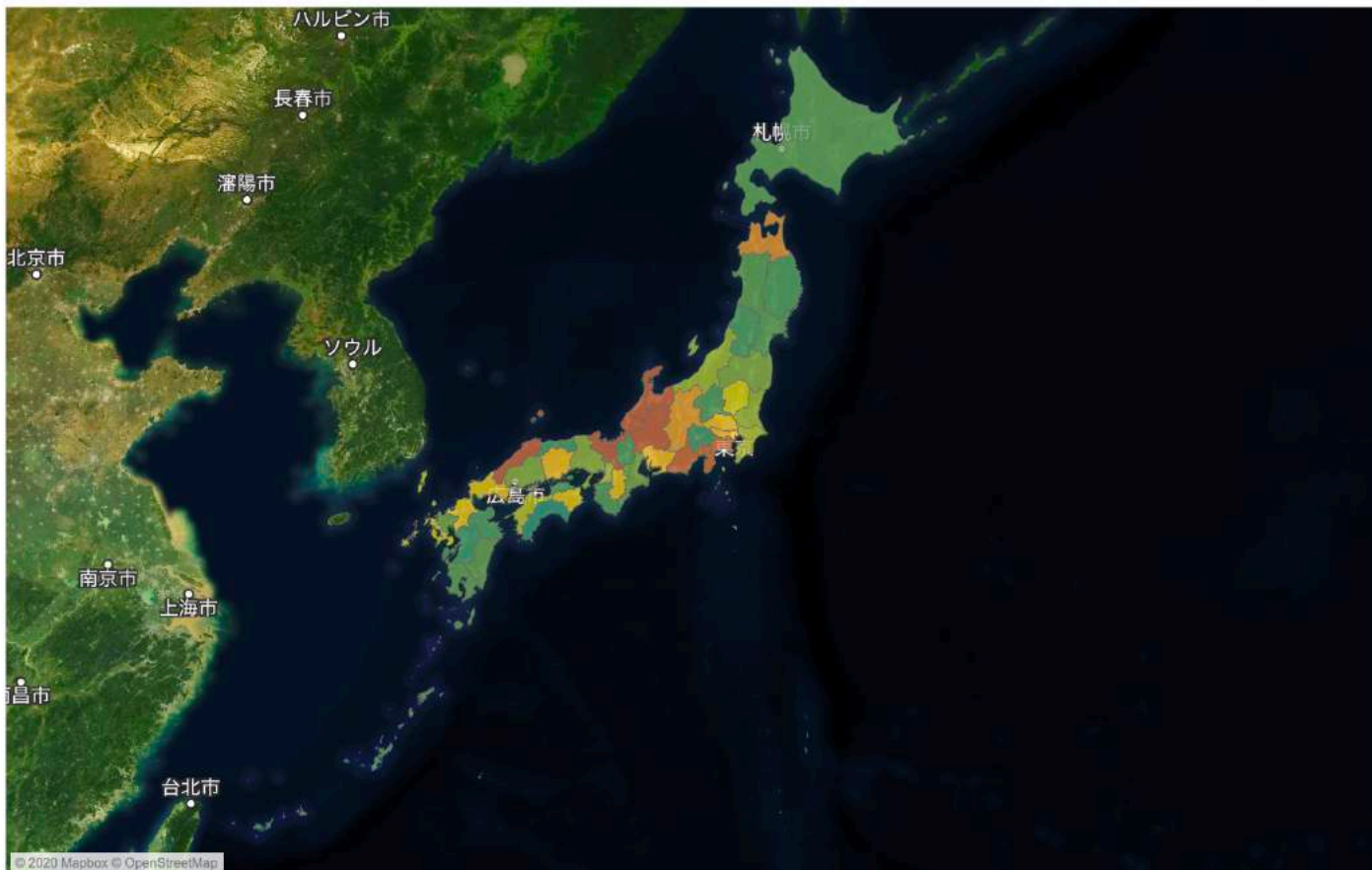
日本のオープンデータ取組自治体

データ出典元：CC BY 4.0 政府CIOポータル>オープンデータ>オープンデータ取組自治体一覧（2020年9月10日時点）

<https://cio.go.jp/policy-opendata#jichitaisuu>

828 / 1,741 自治体 (48%)

地図上の都道府県、または棒グラフをクリックすると、選択した都道府県の詳細が表示されます↓→



都道府県別オープンデータ取組率の比較

1	岐阜県	100%
2	京都府	100%
3	石川県	100%
4	島根県	100%
5	富山県	100%
6	福井県	100%
7	静岡県	97%
8	神奈川県	94%
9	青森県	88%
10	長野県	83%
11	愛知県	74%
12	岡山県	74%
13	東京都	73%
14	埼玉県	71%
15	奈良県	67%

都道府県別オープンデータ取組自治体数の比較

(灰色のバーは自治体数総数)

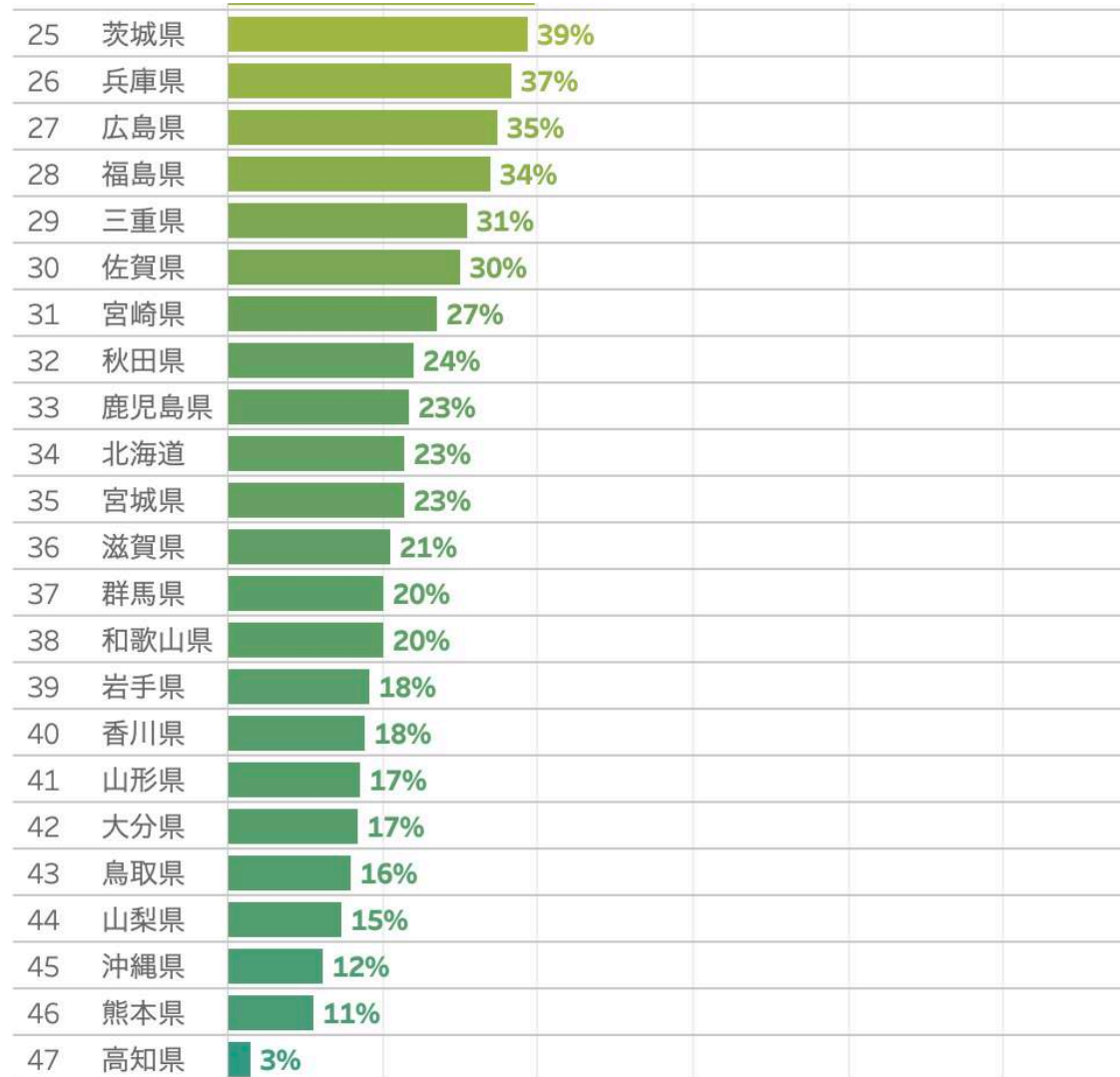
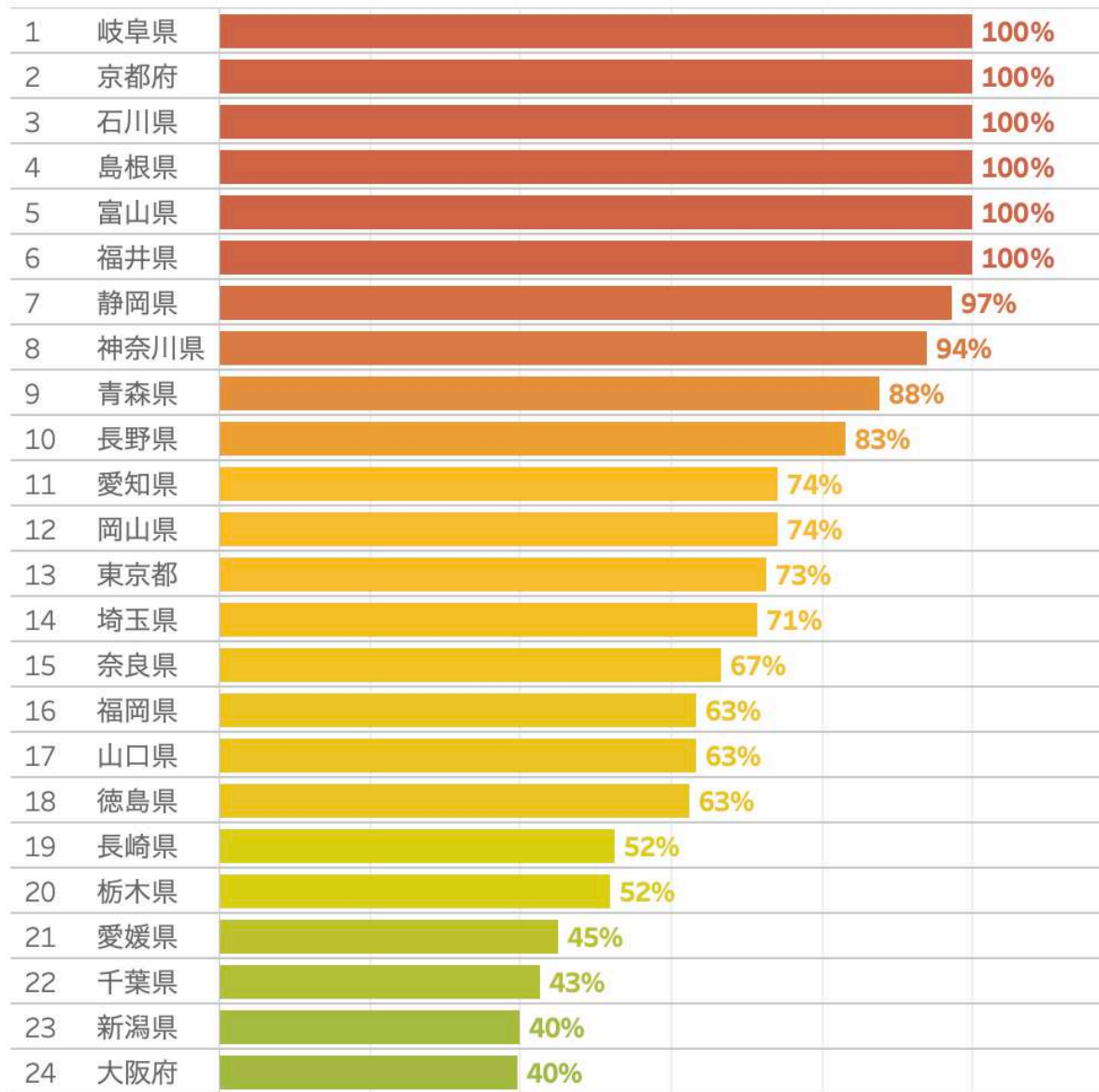
1	長野県	64	77
2	埼玉県	45	63
3	東京都	45	62
4	岐阜県	42	
5	北海道	41	179
6	愛知県	40	54
7	福岡県	38	60
8	青森県	35	0
9	静岡県	34	
10	神奈川県	31	
11	京都府	26	
12	奈良県	26	39
13	千葉県	23	54

オープンデータ取組率



<https://public.tableau.com/profile/sayoko.shimoyama.linkdata#!/vizhome/5946/sheet0>

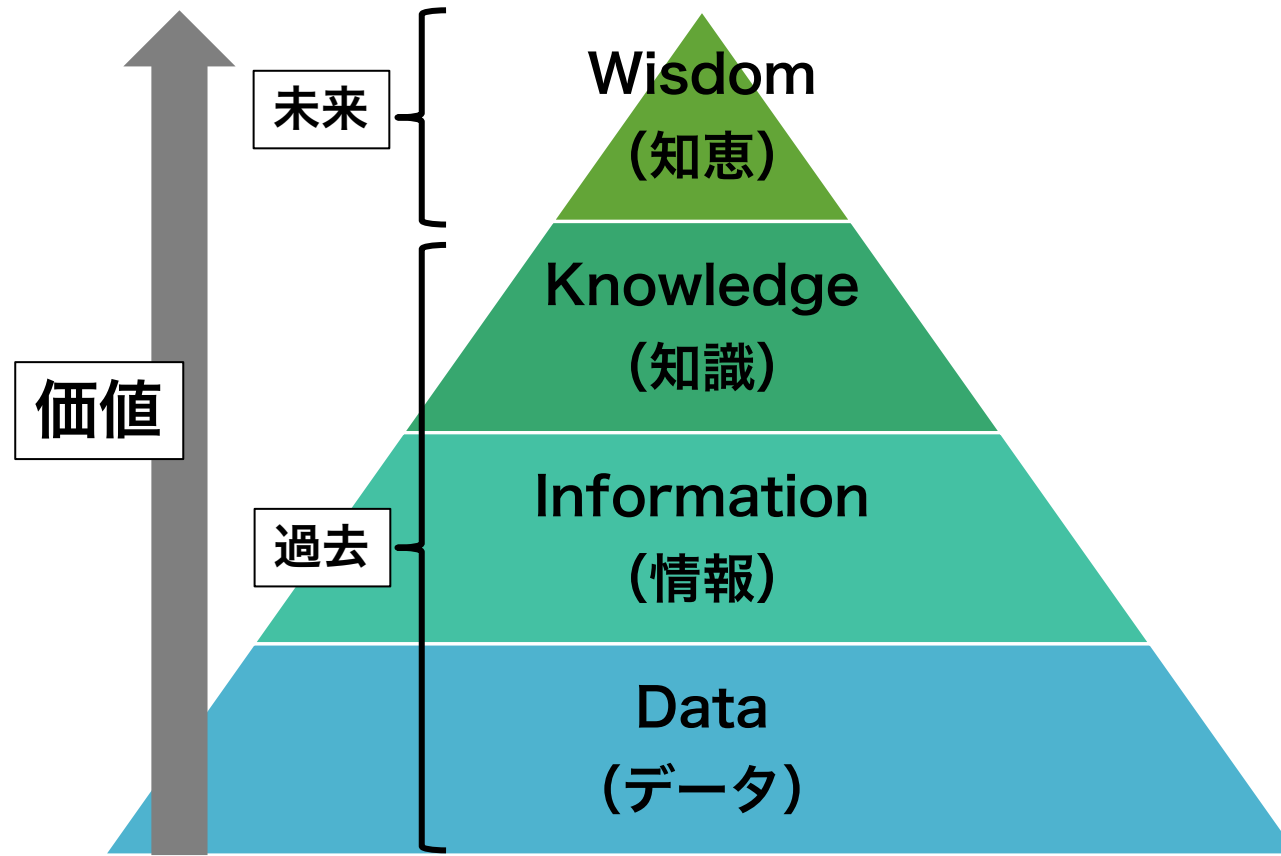
都道府県別オープンデータ取組率の比較



<https://public.tableau.com/profile/sayoko.shimoyama.linkdata#!/vizhome/5946/sheet0>

3. 政策立案のための データ分析の流れ

ポイント： D・I・K・Wのつながりを意識する



知識を活用して判断ができる状態になると「**知恵**」と呼べるものになる。

情報を蓄積し関連付けることで、「**知識**」として事象の規則性や傾向を理解できる。

データ処理によって得られる「**情報**」になってはじめて、意味を持つものになる。

「**データ**」は、単なる（事実や状況を表す）数字や文字列の集まりで、それだけでは意味を持たない。



データ



情報

違いを聞かれたら説明できますか？

「データ」とは？

ISO及びJIS規格による定義

"A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing."

情報の表現であって、
伝達、解釈または処理に適するように形式化され、
再度情報として解釈できるもの

ISO/IEC 2382-1:1993 <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:en>

日本工業規格「X0001 情報処理用語-基本用語」<http://kikakurui.com/x0/X0001-1994-01.html>

「情報」とは？

ISO及びJIS規格による定義

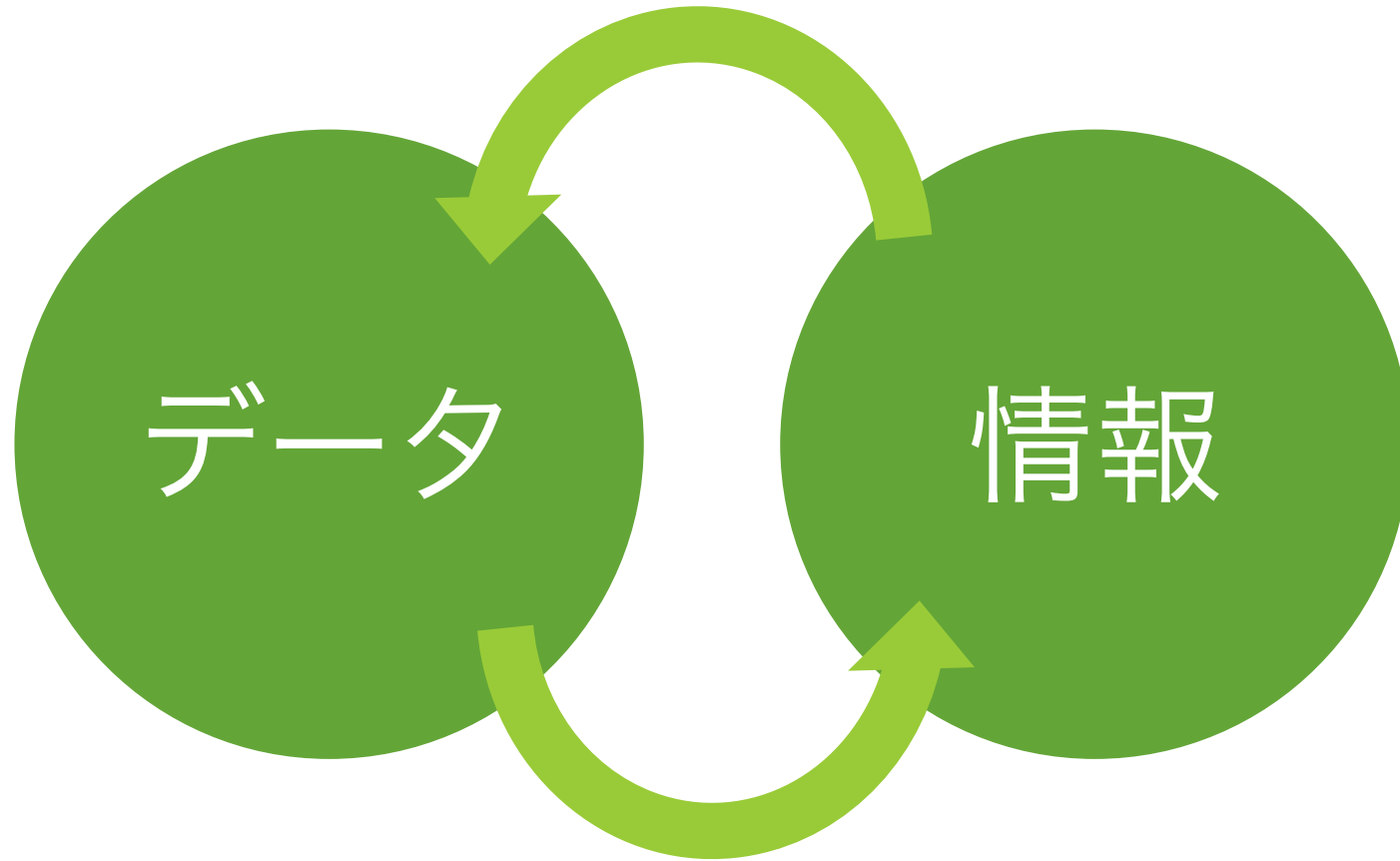
"Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning.

事実，事象，事物，過程，着想などの
対象物に関して知り得たことであって，
概念を含み，一定の文脈中で特定の意味をもつもの。

ISO/IEC 2382-1:1993 <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:en>

日本工業規格「X0001 情報処理用語-基本用語」 <http://kikakurui.com/x0/X0001-1994-01.html>

データは、情報を形式化したもの



再度**情報**として解釈できるもの

40

数字

横浜市役所で
2020年1月23日に開催された
データ分析ワークショップの
参加登録者数は
40人です。

情報

ワークショップ 開催日	ワークショップ 開催場所	参加登録者 数
2020-1-23	横浜市役所	40

データ

情報⇔データ 双方向に変換可能

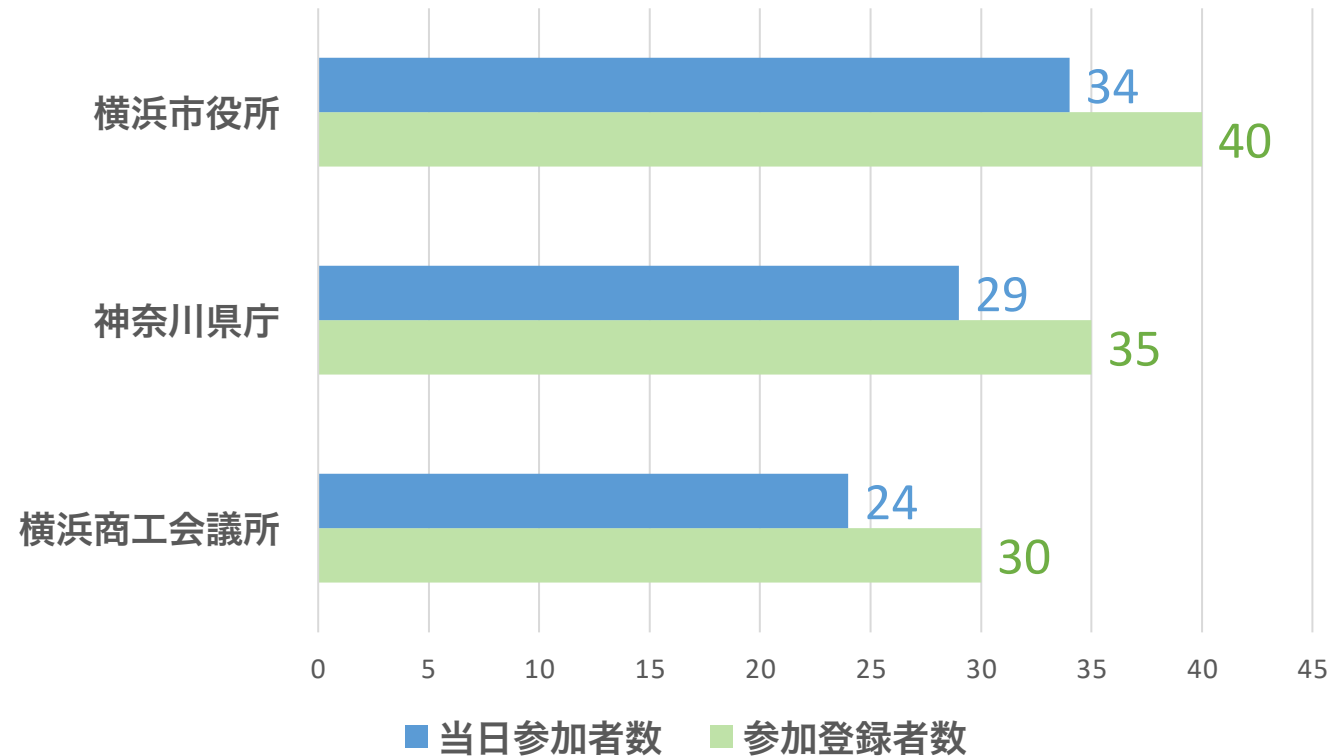
横浜市役所で
2020年1月23日に開催された
データ分析ワークショップの
参加登録者数は
40人です。



ワークショップ 開催日	ワークショップ 開催場所	参加登録 者数
2020-1-23	横浜市役所	40

データ化することで 他のデータと組み合わせられる

ワークショップ別 参加者数の比較



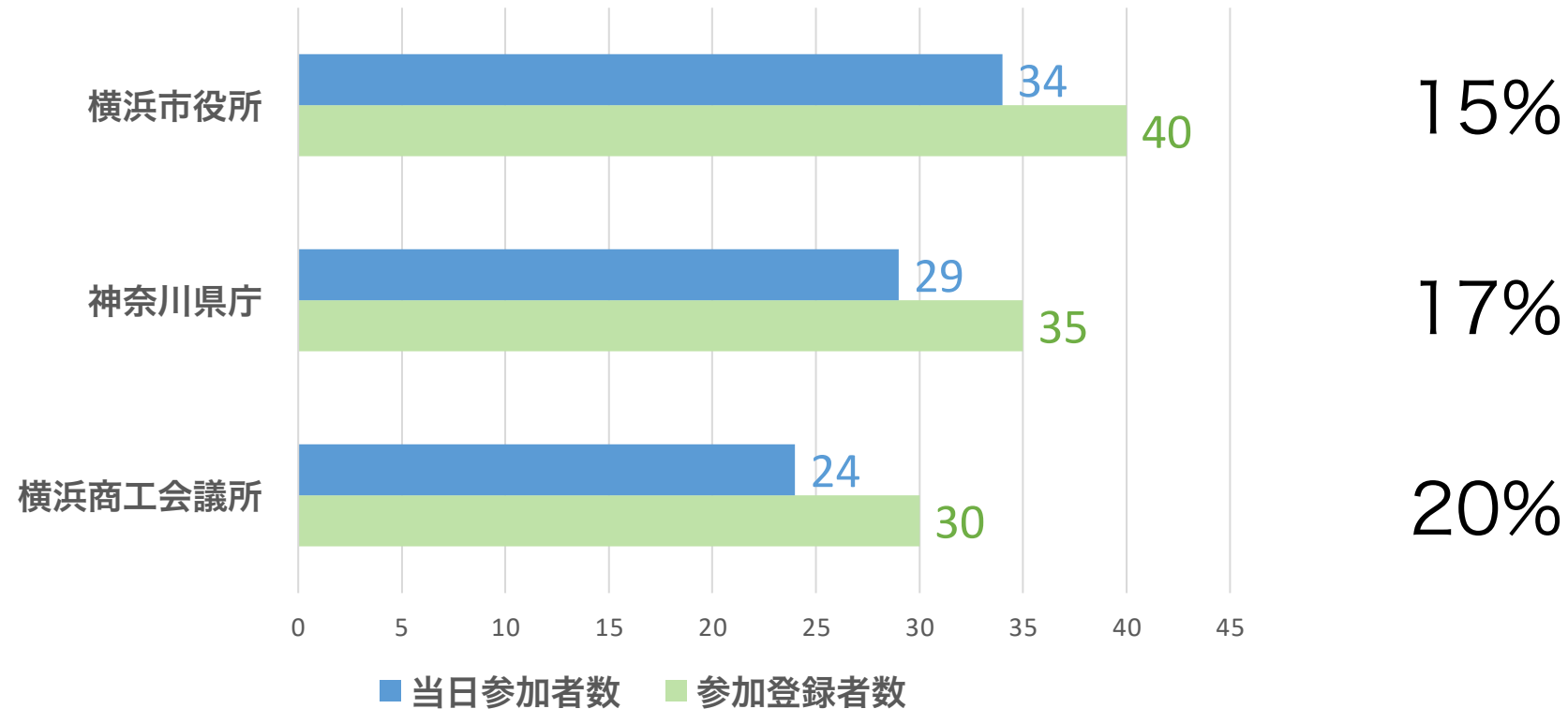
新しい「情報」が
取り出せる

Q：各ワークショップの
「キャンセル率」は
どのくらい？

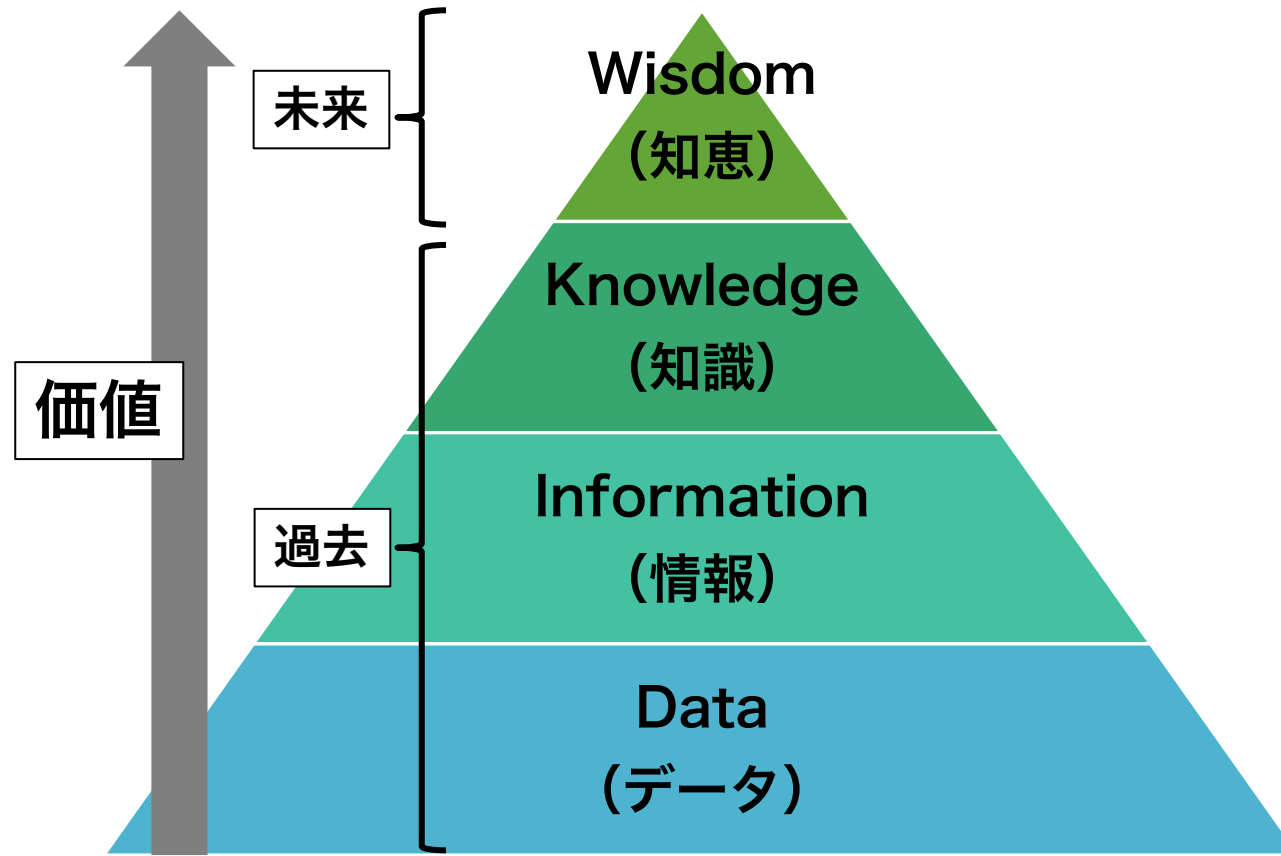
データ分析 = データから情報を引き出すこと

ワークショップ別 参加者数の比較

キャンセル率



DIKWモデルに当てはめた例



次回から定員+15~20%くらい
多めに申込を受け付けても
大丈夫そう

ワークショップのキャンセル率
は大体毎回15%~20%くらい

各ワークショップの
キャンセル率

各ワークショップの
登録者数と当日参加者数

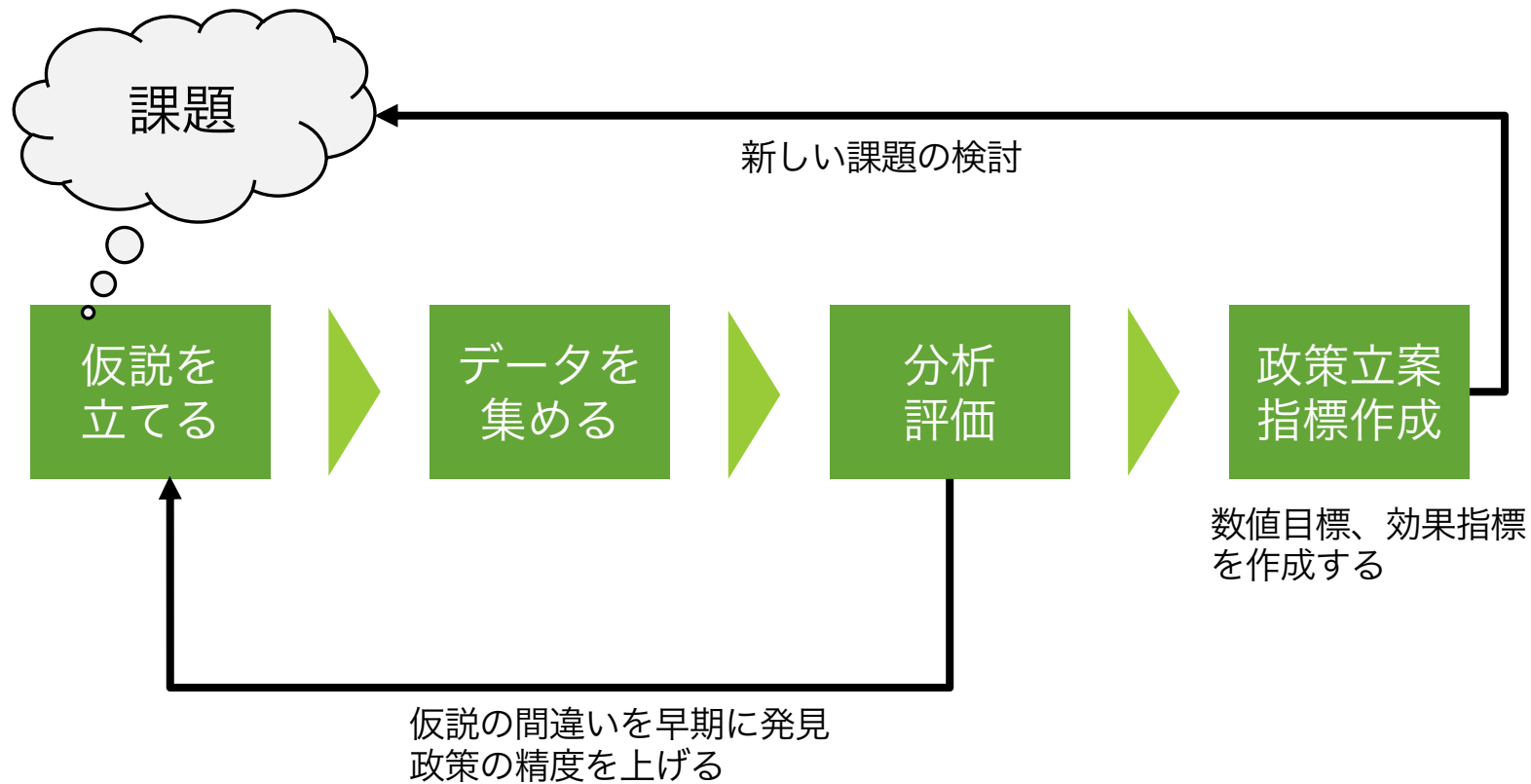
注意：「知恵」は恒久的なものではない

- 現地でイベントを行っていた時とオンラインイベントではキャンセル率は変わるのでは？
- 参加者層が全然違うイベントの場合は？



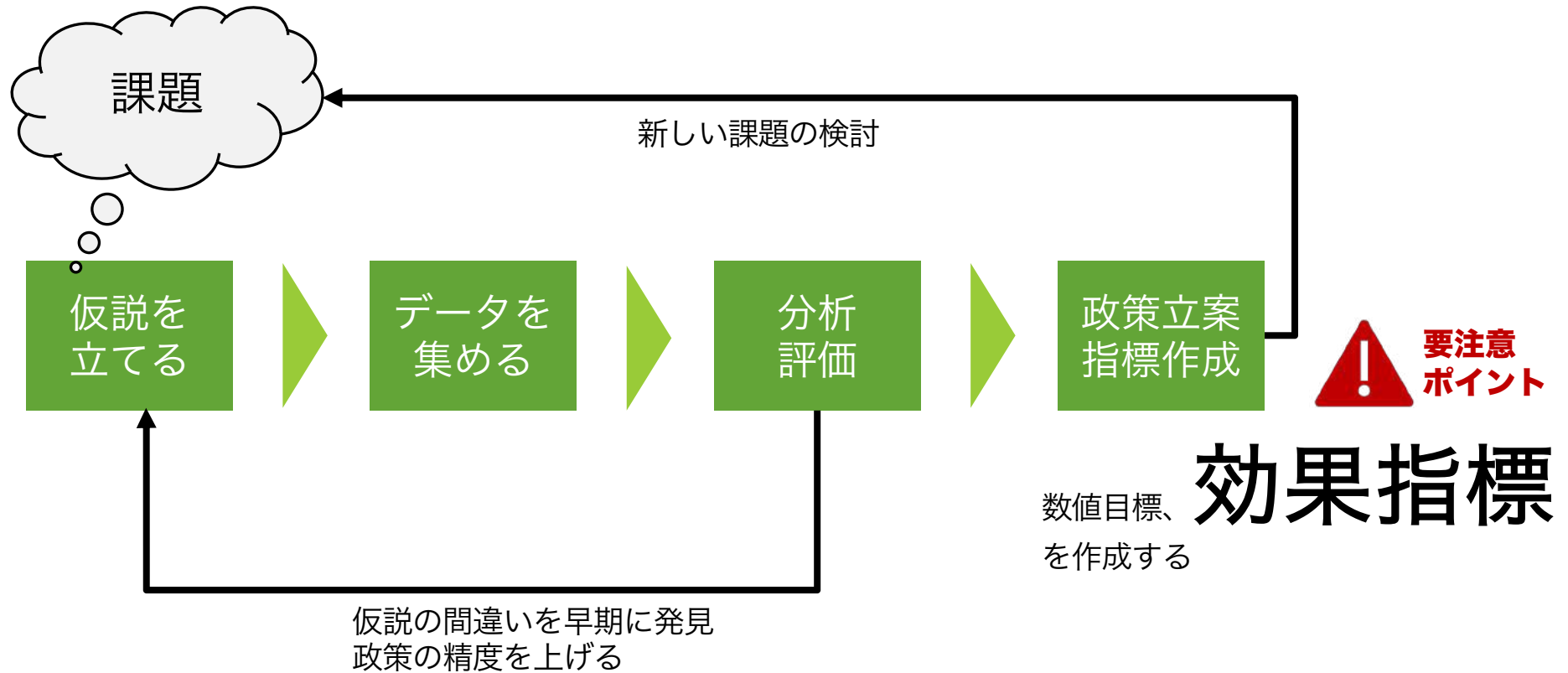
変化に対応して「知恵」をアップデートするためには、
DIKWを行ったり来たりする必要がある

データ分析による政策反映の流れ



出典元：総務省 地方公共団体におけるデータ活用ガイドブック Ver. 2.0
別添資料4. データアカデミー（データ分析編）の教材・資料等, CC BY 2.1 JP

データ分析による政策反映の流れ



出典元：総務省 地方公共団体におけるデータ活用ガイドブック Ver. 2.0
別添資料4. データアカデミー（データ分析編）の教材・資料等, CC BY 2.1 JP

効果指標：定性的効果をどう測る？

住みやすさ

まちの
にぎわい

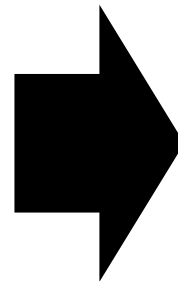
老後の安心

子育ての
しやすさ

男女平等



アンケートで聞く
という方法がよく取られる



指標の1つにするのはOK



アンケート結果のみで測ると
客観性に欠ける場合がある

ポイント：
定性的効果は
定量的効果に
分解して
指標化する

例：住みよさランキング2020
(東洋経済)

■「住みよさランキング」総合評価トップ50(1~25位)

順位	都市名(都道府県名)	総合評価 偏差値	順位			
			安心度	利便度	快適度	富裕度
1	白山(石川)	54.231	116	308	8	77
2	文京区(東京)	54.099	535	9	53	68
3	野々市(石川)	54.052	121	11	182	488
4	福井(福井)	54.039	165	42	213	67
5	倉吉(鳥取)	53.873	5	60	276	654
6	敦賀(福井)	53.756	104	110	154	130
7	黒部(富山)	53.734	150	626	5	35
8	能美(石川)	53.730	84	632	11	70
9	魚津(富山)	53.687	192	24	339	92
10	駒ヶ根(長野)	53.652	33	92	209	376
11	四万十(高知)	53.625	46	18	22	775
12	郡上(岐阜)	53.462	37	227	114	317
13	かほく(石川)	53.332	127	117	52	290
14	印西(千葉)	53.224	205	161	229	37
15	七尾(石川)	53.153	16	89	582	332
16	小松(石川)	53.144	125	127	306	103
17	北杜(山梨)	53.140	201	105	7	413
18	渋谷区(東京)	53.129	798	1	117	71
19	合志(熊本)	53.096	29	804	15	181
20	金沢(石川)	53.077	300	32	175	184
21	にかほ(秋田)	53.058	25	310	387	223
22	下松(山口)	52.988	369	106	6	204
23	大阪(大阪)	52.931	796	7	2	111
24	新宿区(東京)	52.893	790	3	70	71
25	武蔵野(東京)	52.884	793	4	104	62

TOYOKEIZAI ONLINE

定性的効果は定量的効果に分解して指標化する

例：住みよさランキング2020（東洋経済）利用データ

住みよさ	A. 安心度	1 人口当たり病院・一般診療所病床数（2018年10月）：厚生労働省「医療施設調査」
		2 老年人口当たり介護老人福祉・保健施設定員数（2017年10月）：厚生労働省「介護サービス施設・事業所調査」
		3 20～39歳女性人口当たり0～4歳児数（2019年1月）：総務省「住民基本台帳に基づく人口、人口動態及び世帯数」
		4 子ども医療費助成（対象年齢・所得制限の有無）（2020年4月）：東洋経済調べ
		5 人口当たり刑法犯認知件数（2018年）：各都道府県警察調べ
		6 人口当たり交通事故件数（2018年）：交通事故総合分析センター調べ
	B. 利便度	7 人口当たり小売販売額（2015年）：総務省・経済産業省「経済センサス活動調査」
		8 人口当たり大規模小売店店舗面積（2019年）：東洋経済「全国大型小売店総覧」
		9 可住地面積当たり飲食料品小売事業所数（2016年6月）：総務省・経済産業省「経済センサス活動調査」
		10 人口当たり飲食店数（2016年6月）総務省・経済産業省「経済センサス活動調査」
	C. 快適度	11 転出入人口比率（2018年）：総務省「住民基本台帳に基づく人口、人口動態及び世帯数」
		12 水道料金（2020年4月）：東洋経済調べ
		13 汚水処理人口普及率（2019年3月）：国土交通省、農林水産省、環境省調べ
		14 都市計画区域人口当たり都市公園面積（2018年3月）：国土交通省「都市公園整備水準調書」
		15 気候（月平均最高・最低気温、日照時間、最深積雪【新】）（1981～2010年）：気象庁「メッシュ平年値データ」
	D. 富裕度	16 財政力指数（2018年度）：総務省「市町村別決算状況調」
		17 人口当たり法人市民税（2018年度）：総務省「市町村別決算状況調」【新】
		18 納税義務者1人当たり所得（2018年）：総務省「市町村税課税状況等の調」
		19 1住宅当たり延べ床面積（2018年10月）：総務省「住宅・土地統計調査」
		20 住宅地平均地価（2019年7月）：国土交通省「都道府県地価調査」

<https://toyokeizai.net/articles/-/356816?page=2>

指標は定期的に見直して改善することも重要

例：住みよさランキング2019（東洋経済）利用データ

住みよさ	A. 安心度	1 人口当たり病院・一般診療所病床数（2017年10月）：厚生労働省「医療施設調査」
		2 老年人口当たり介護老人福祉・保健施設定員数（2017年10月）：厚生労働省「介護サービス施設・事業所調査」
		3 20～39歳女性人口当たり0～4歳児数（2018年1月）：総務省「住民基本台帳に基づく人口、人口動態及び世帯数」
		4 子ども医療費助成（対象年齢・所得制限の有無）（2019年6月）：東洋経済調べ
		5 人口当たり刑法犯認知件数（2017年）：各都道府県警察調べ
		6 人口当たり交通事故件数（2017年）：交通事故総合分析センター調べ
	B. 利便度	7 人口当たり小売販売額（2015年）：総務省・経済産業省「経済センサス活動調査」
		8 人口当たり大規模小売店店舗面積（2018年）：東洋経済「全国大型小売店総覧」
		9 可住地面積当たり飲食料品小売事業所数（2016年6月）：総務省・経済産業省「経済センサス活動調査」
		10 人口当たり飲食店数（2016年6月）総務省・経済産業省「経済センサス活動調査」
	C. 快適度	11 転出入人口比率（2015～2017年）：総務省「住民基本台帳に基づく人口、人口動態及び世帯数」
		12 人口当たり財政歳出額（2017年度）：総務省「市町村別決算状況調」
		13 水道料金（2019年6月）：東洋経済調べ
		14 汚水処理人口普及率（2018年3月）：国土交通省、農林水産省、環境省調べ
		15 都市計画区域人口当たり都市公園面積（2017年3月）：国土交通省「都市公園整備水準調書」
		16 気候（月平均最高・最低気温、日照時間）（1981～2010年）：気象庁「メッシュ平年値データ」
	D. 富裕度	17 財政力指数（2017年度）：総務省「市町村別決算状況調」
		18 1事業所当たり売上高（2016年6月）：総務省・経済産業省「経済センサス活動調査」
		19 納税義務者1人当たり所得（2017年）：総務省「市町村税課税状況等の調」
		20 1住宅当たり延べ床面積（2013年10月）：総務省「住宅・土地統計調査」
		21 持ち家世帯比率（2015年10月）：総務省「国勢調査」
		22 住宅地平均地価（2018年7月）：国土交通省「都道府県地価調査」

SDGsでもグローバルな目標達成のために 指標が使われている

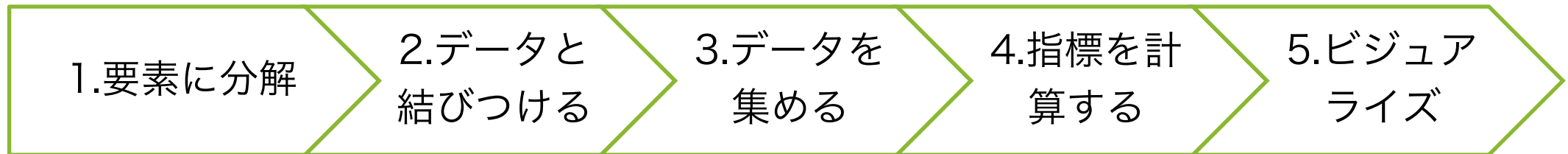


参考： Sustainable Development Report による ダッシュボード

Cambridge University Press,
CC BY 4.0
<https://dashboards.sdgindex.org/profiles/JPN>



ワークショップ形式による 指標作成のステップ



構成要素として考えられるものを列挙してツリー構造になるように整理する

それぞれの構成要素を測るためにどのようなデータが必要か考える

2. で検討したデータが実際にあるか探して収集する

収集したデータを用いて、比較対象別に指標を計算する

算出した指標値を比較しやすくするために可視化する

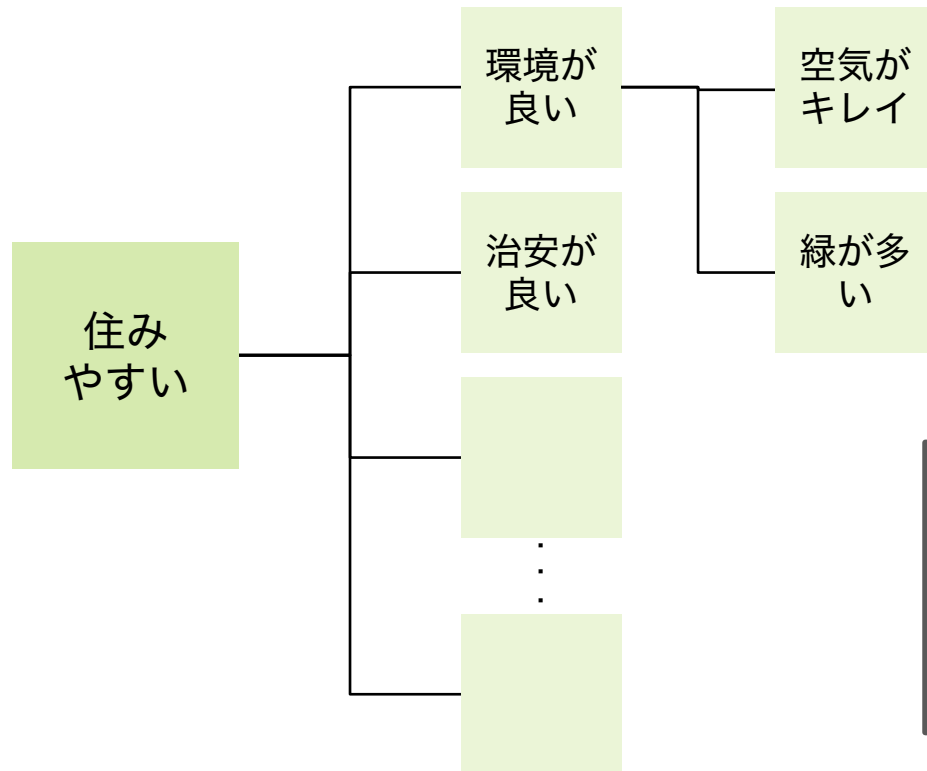
子育て世代にとっての “住みやすさ”

の指標を作成し、
周辺自治体と定量的に
比較可能にする



Step 1. 要素に分解

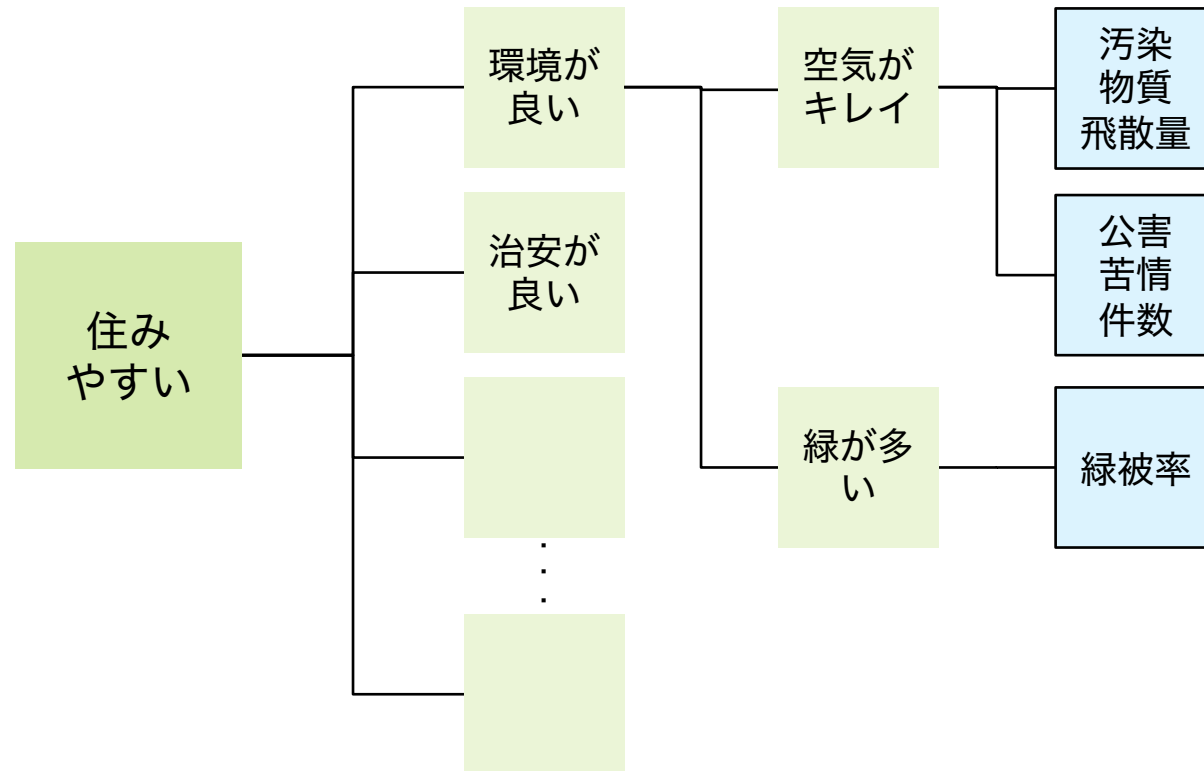
子育て世代にとっての「住みやすい」という状況が成り立つには何が
必要か、挙げてみましょう



- 「これが正解」というものは無いので、思いつくものをどんどん挙げてみましょう
- ここでは一旦データがあるかどうかは気にしなくてOKです

Step2. データと結びつける

Step1で挙げた各要素は、どんなデータで測ることができそうか、挙げてみましょう



Step3. データを集める

ヒント：データが無い場合、他のデータで代替や近似ができないか考えてみましょう

- Step2で挙げたデータが実際にあるか探してみましよう
- データを探す中で、他にも指標として使えそうなものを見つけたら、追加してもOKです

Step4. 指標を計算する

1. 指標に使うデータの種類ごとに各区の **偏差値** を計算する
2. 各地域の **総合偏差値** を計算する

(計算のイメージ)

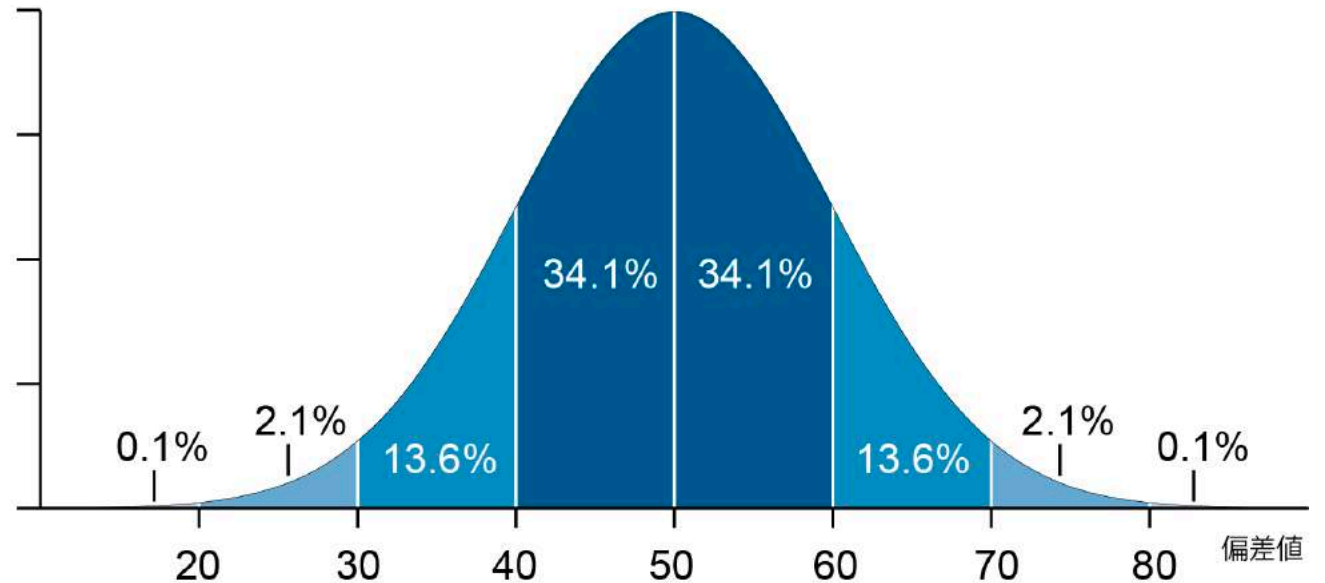
	偏差値_緑被率	偏差値_汚染物質	偏差値_XXX	総合偏差値
A市	41.23	55.93	...	52.55
B市	34.24	37.71	...	34.89
C市	58.13	35.03	...	53.76
...

参考：偏差値を利用する理由

- 指標に使うデータ毎の性質の違いを平準化し、比較可能にするため

偏差値とは

- ある数値が集団の平均値からどのくらい離れているかを表す指標
- 平均値やばらつき具合が異なるデータの集合を比較可能にするために、データを平均が50、標準偏差が10となるように規格化した値



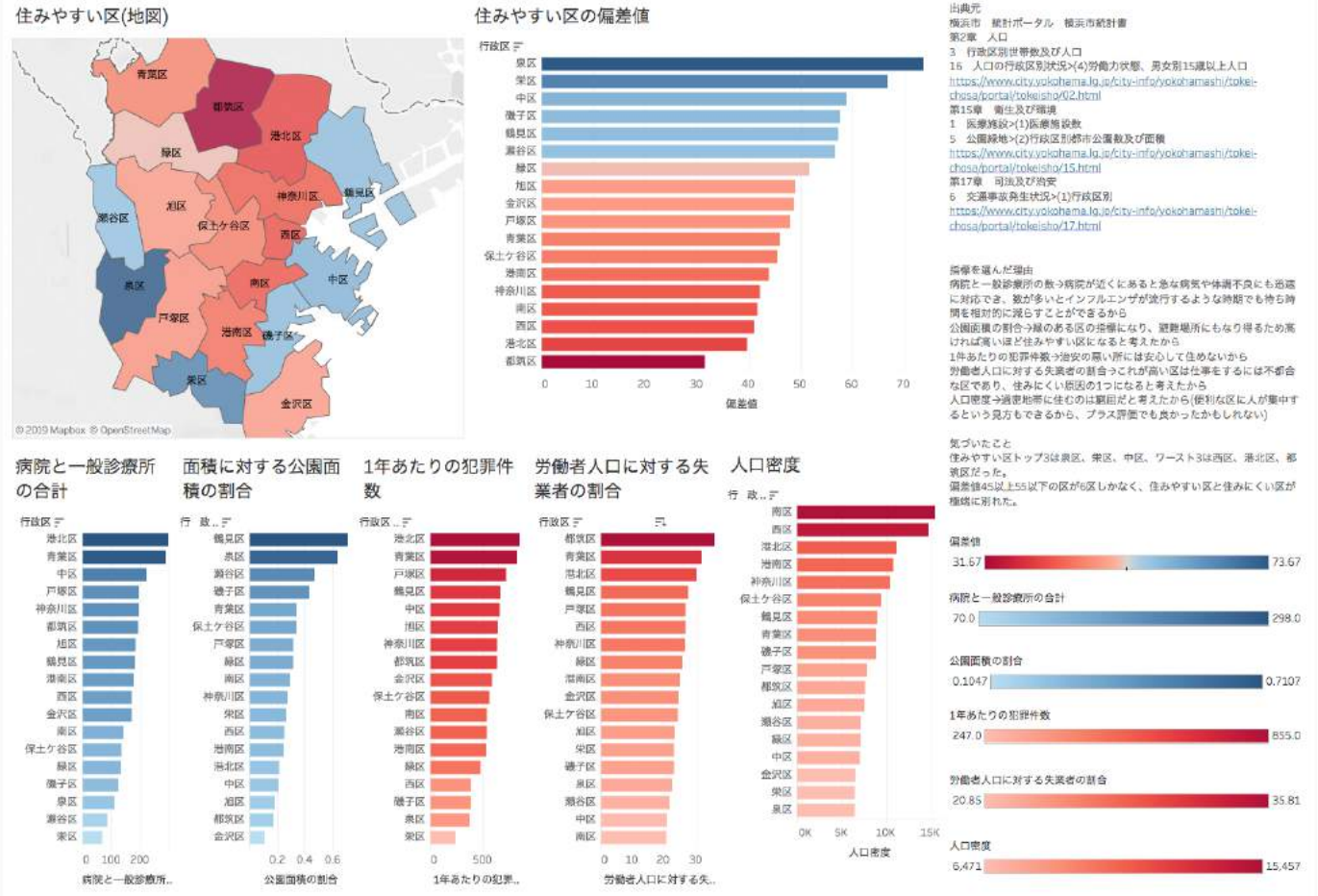
偏差値が大きいほど「住みやすさ」の評価が高い

Step5. ビジューアライズ

算出した指標値を比較しやすくするために可視化する

※横浜情報科学専門学校の学生チームによる「横浜市18区の住みやすさ」指標の可視化例

<https://public.tableau.com/profile/.48315447#!/vizhome/7877/sheet7>



指標設計体験ワーク

次のうち一番興味のあるテーマを選んでください

住みやすさ

まちの
にぎわい

老後の安心

子育ての
しやすさ

男女平等

指標設計体験ワーク

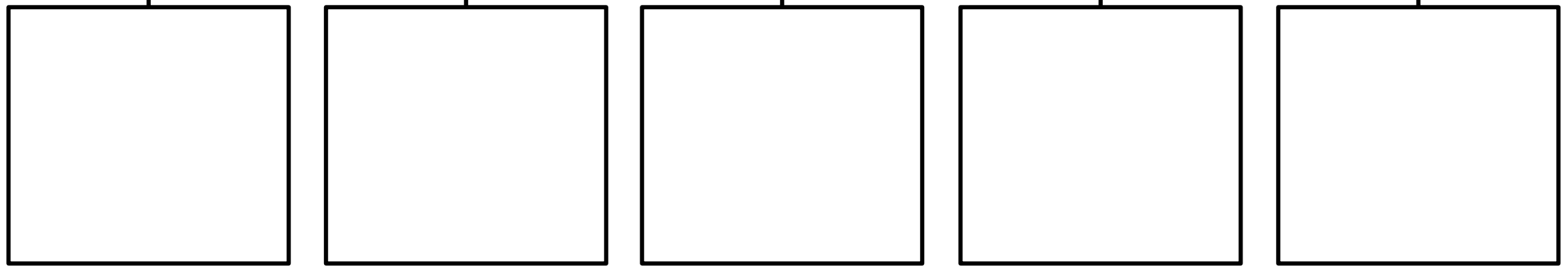
テーマ

※今回はStep2まで

Step1 : 要素に分解



Step2 : データと結びつける

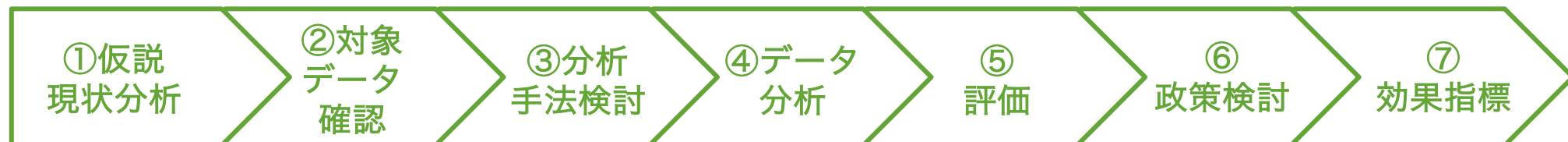


データ分析の流れ全体を実際に体験したい方にオススメ

データアカデミー



- ITで地域課題の解決を進める団体「Code for Japan」が実施する、行政職員向けデータ利活用研修プログラム
- 2017年度の総務省事業においてカリキュラム開発を実施
- これまでに60以上の自治体が参加
- 従来の座学による研修とは異なり、職員が普段の業務で使用しているデータを用いて、政策の立案や検証をおこなうことが最大の特徴
- データ分析による政策反映の7stepの流れを、半日×4日間の研修で習得する



詳細はこちら→データアカデミー研究会 <https://da.code4japan.org/>

4. データ活用における よくある勘違い

データ活用におけるよくある勘違い

1. 定性データは役に立たない
2. 相関があれば必ず因果関係がある
3. エビデンスを求められた時は、データを示せばOK!

データ活用における よくある勘違い

1. 定性データは役に立たない
2. 相関があれば必ず因果関係がある
3. エビデンスを求められた時は、
データを示せばOK!

データは性質で大きく2つに分けられる

定量データ

- 数値で大小を表すことができるデータ
- 客観性が高い
(誰が見ても大体同じように解釈できる)
- マクロな視点での分析に用いる
(鳥の目)

定性データ

- 数値で表現できない文章などのデータ
- 主観性が高い
(解釈はその人の持つ尺度に委ねられる)
- ミクロな視点での分析に用いる
(虫の目)

定量データの例： アンケートの選択項目

某市における行政職員向けデータ活用研修のアンケートより

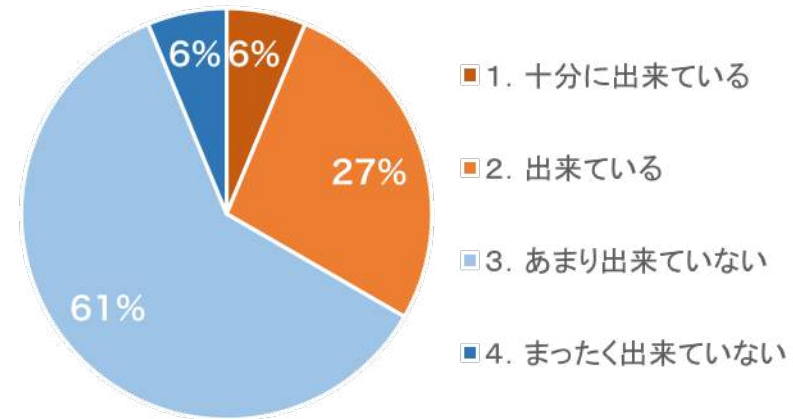
あなたは、日々の業務の中で、データを活用出来ていると思いますか？

1. 十分に出来ている
2. 出来ている
3. あまり出来ていない
4. まったく出来ていない



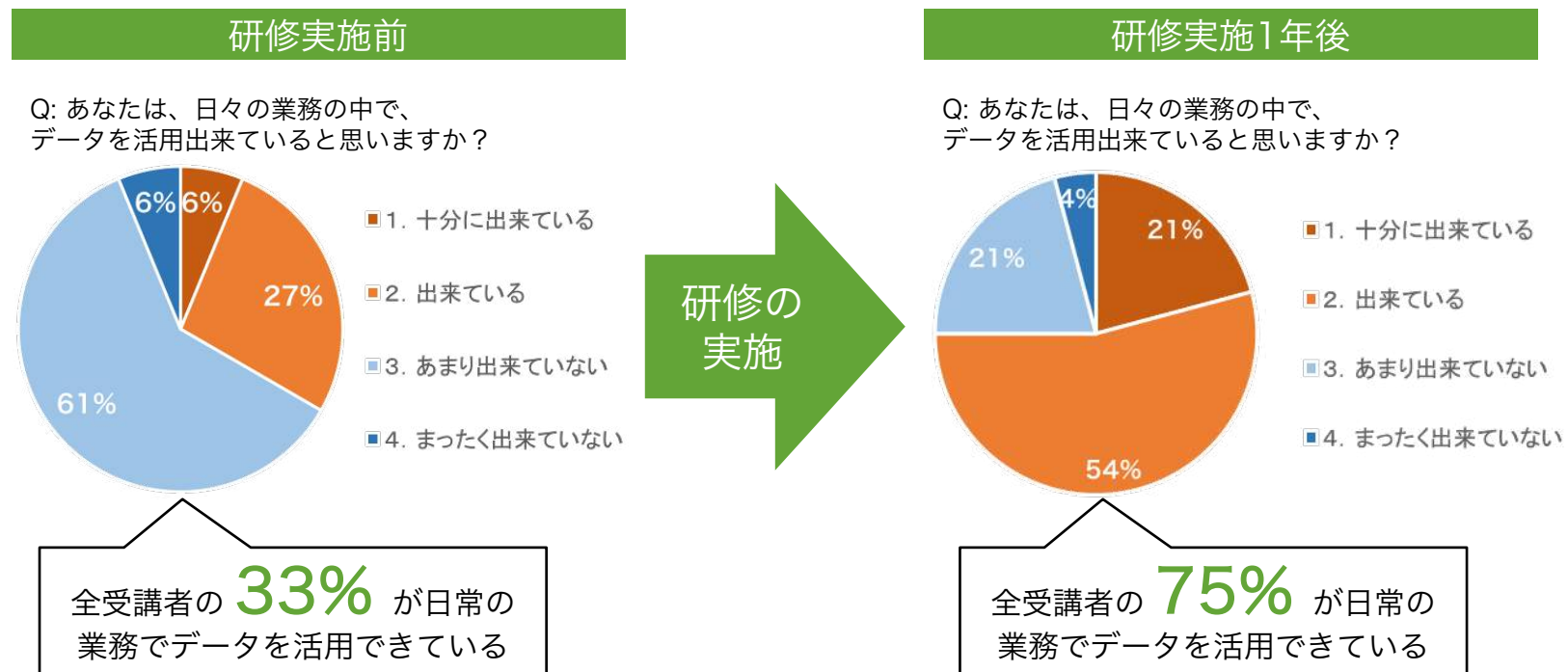
集計して数値化できる = **定量データ**

1. 十分に出来ている	3
2. 出来ている	13
3. あまり出来ていない	29
4. まったく出来ていない	3



定量データの使いどころは「俯瞰」

- 数値で比較することで、状態を俯瞰的に把握できる
- 統一された基準で比較できる



定性データの例： アンケートの自由記述項目

某市における行政職員向けデータ活用研修のアンケートより

データ活用について、課題等があれば教えてください。（自由記述）



数値化できない
= 定性データ

データ収集、活用、提供のノウハウ不足
現在の業務内容上、あまり多くのデータを扱わないため、活用方法に悩んでいる
オープンデータ概念が全く浸透していない、仕事もデータもタテ割り
PC等、ハード・ソフト環境が不足している（処理能力等）
普段の業務から活用した後のイメージがわからない
データを見て何かを感じとる、把握することに苦手意識があり活用まで至らない、センス？慣れ？
未知の有効活用されていない庁内データが存在すると思われる
...

定性データの使いどころは「深掘り」

- 定量データよりも情報量は多いため、個々の状態をより深掘りして把握できる
- 次のアクションを検討する上での示唆が得られる

Q: データ活用について、課題等があれば教えてください。

データ収集、活用、提供のノウハウ不足
現在の業務内容上、あまり多くのデータを扱わないため、活用方法に悩んでいる
オープンデータ概念が全く浸透していない、仕事もデータもタテ割り
PC等、ハード・ソフト環境が不足している（処理能力等）
普段の業務から活用した後のイメージがわからない
データを見て何かを感じとる、把握することに苦手意識があり活用まで至らない、センス？慣れ？
未知の有効活用されていない社内データが存在すると思われる
...



- データを自由に使える環境が整っていない
- 自分の業務と結び付けて考えることが困難
- データを使うことに対する漠然とした苦手感覚がある



個別のインタビュー等でより状況を深く把握

データ活用における よくある勘違い

1. ~~定性データは役に立たない~~
→ 定量データと定性データは目的に応じて使い分ける
2. 相関があれば必ず因果関係がある
3. エビデンスを求められた時は、
データを示せばOK!

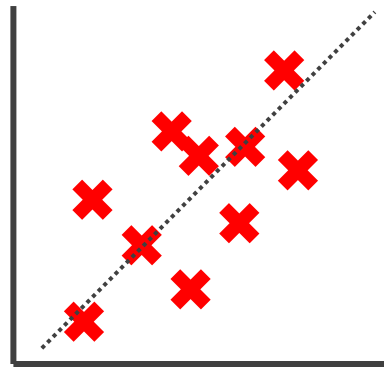
データ活用におけるよくある勘違い

1. ~~定性データは役に立たない~~
→ 定量データと定性データは目的に応じて使い分ける
2. 相関があれば必ず因果関係がある
3. エビデンスを求められた時は、
データを示せばOK!

相関とは

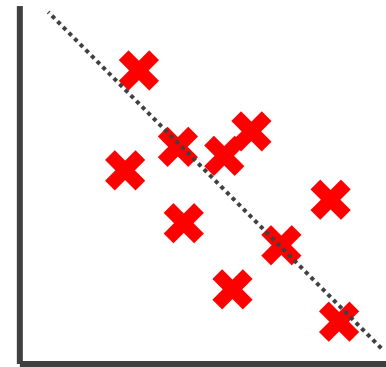
二つの事象が密接にかかわり合い、一方が変化すれば他方も変化するような関係。

正の相関



一方の値が大きくなると、
もう一方の値も大きくなる

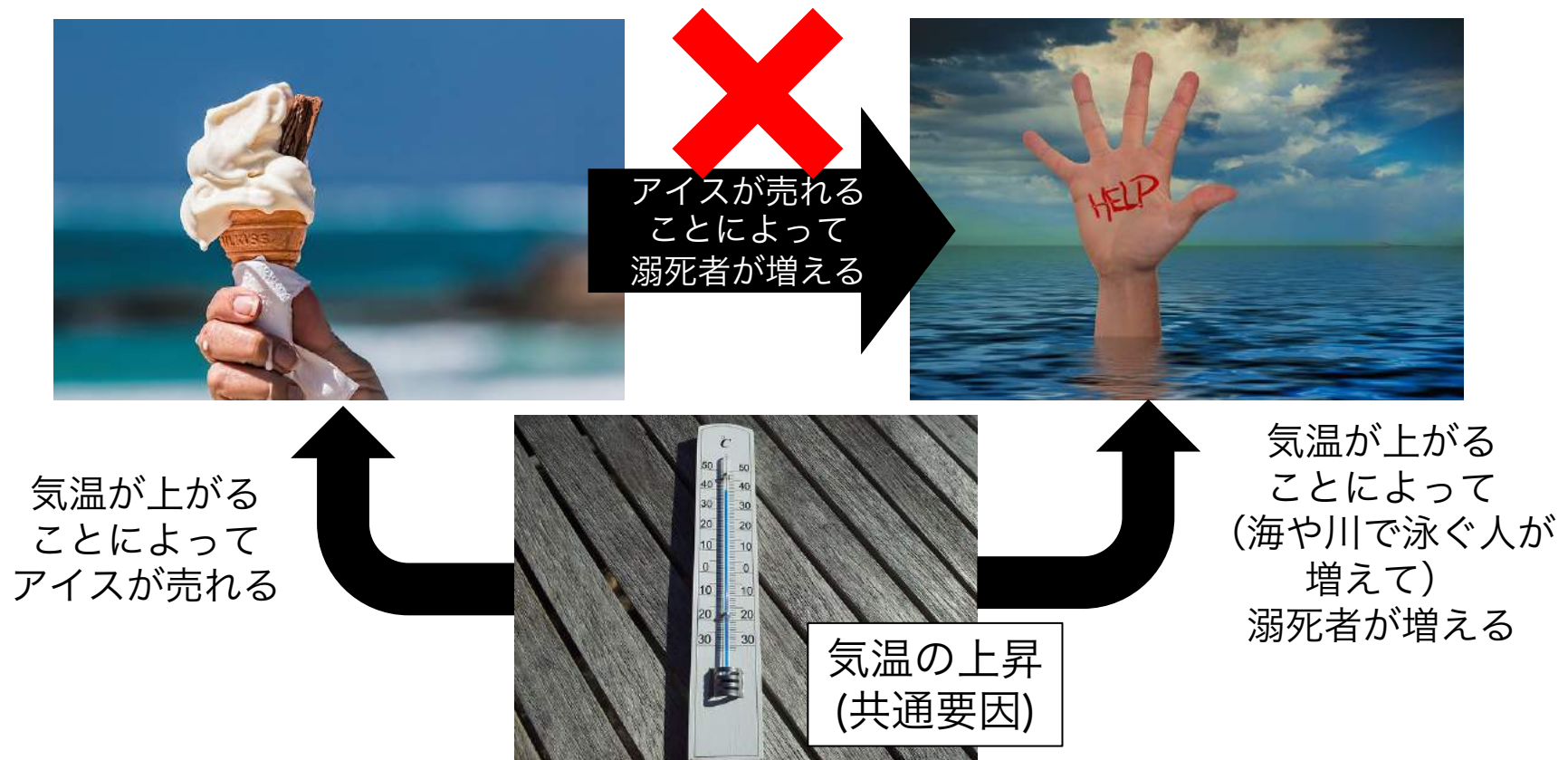
負の相関



一方の値が大きくなると、
もう一方の値は小さくなる

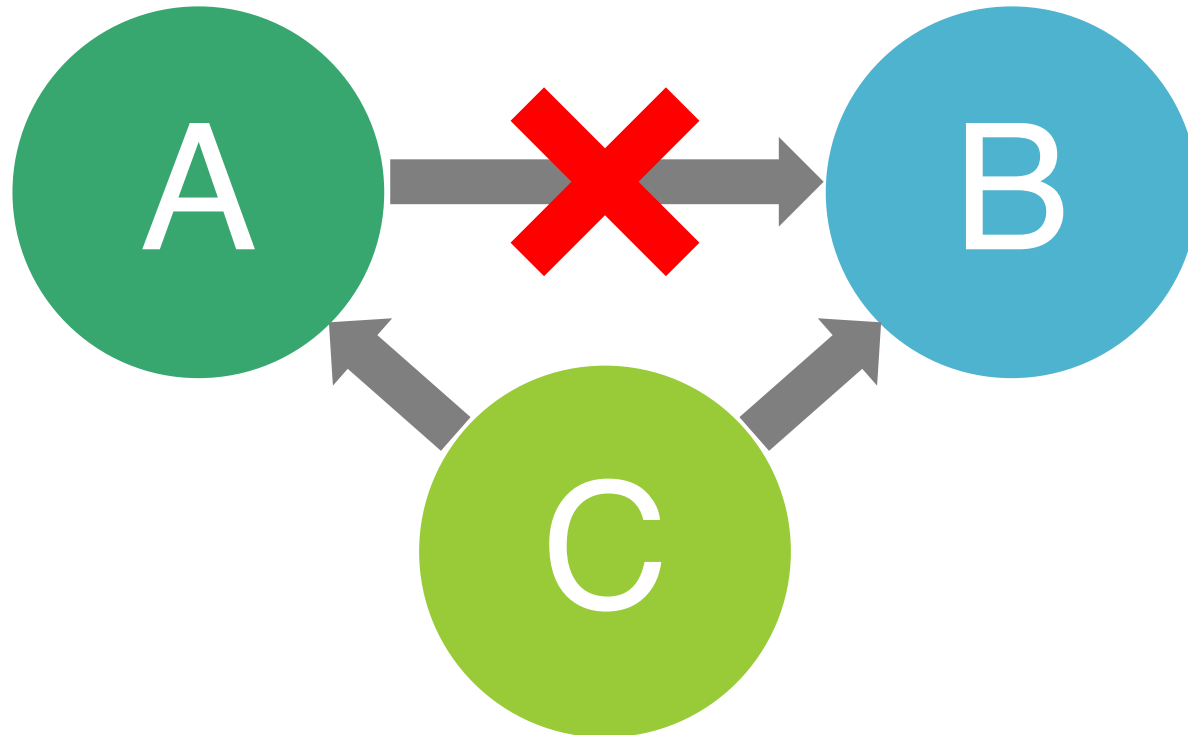
相関≠因果関係

例) アイスクリームの売り上げと溺死者の数に相関があった



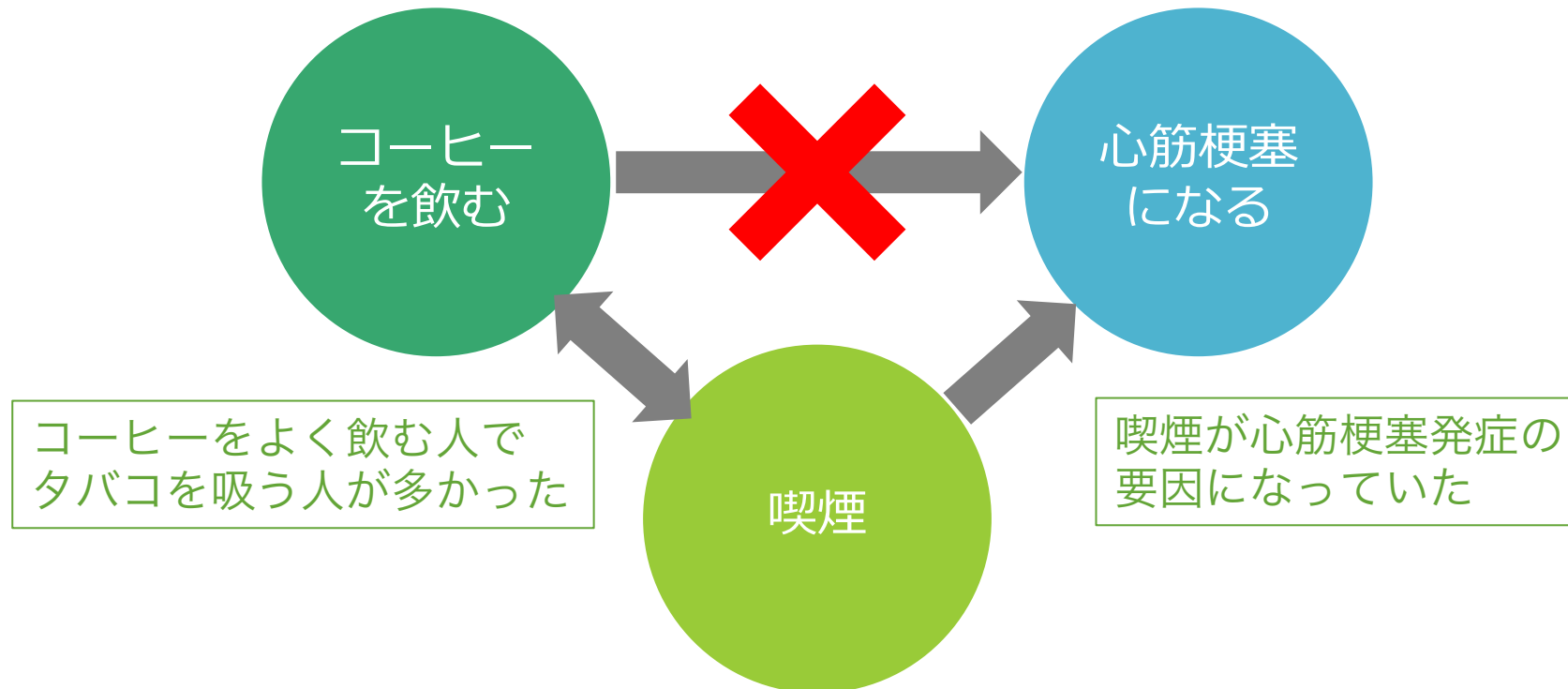
相関があっても因果関係がないケース①： 共通の要因がある（交絡）

- CがAとBに影響を与えている状態
- AとBの間に直接的な関係性はない



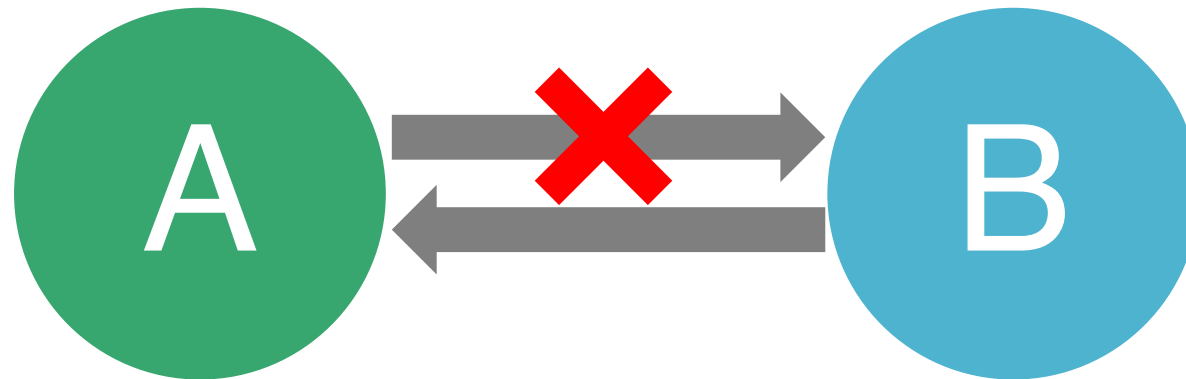
相関があっても因果関係がないケース①： 共通の要因がある（交絡）

- 例) 「コーヒーを飲む量」と「心筋梗塞の発症率」に正の相関があった



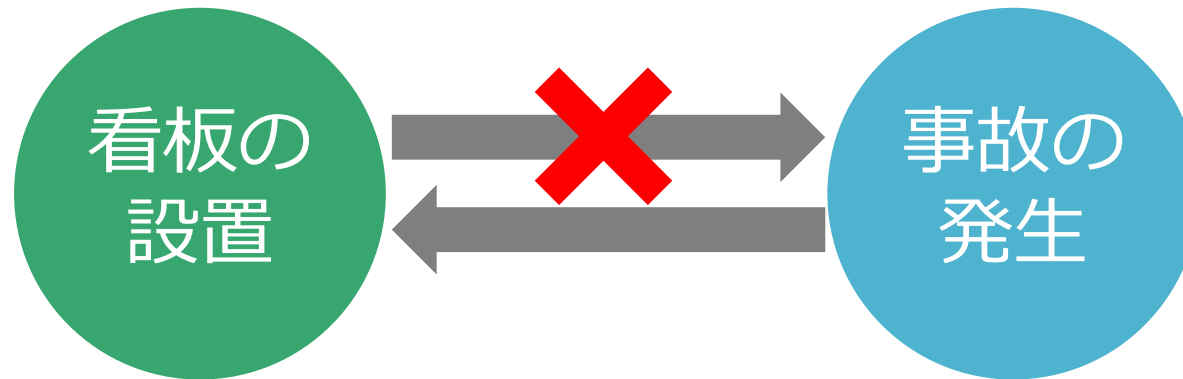
相関があっても因果関係がないケース②： 因果関係の方向が逆

- AによってBが引き起こされたのではなく、BによってAが引き起こされた
- 相関分析では時間的な前後関係が見えないため、気づきにくい



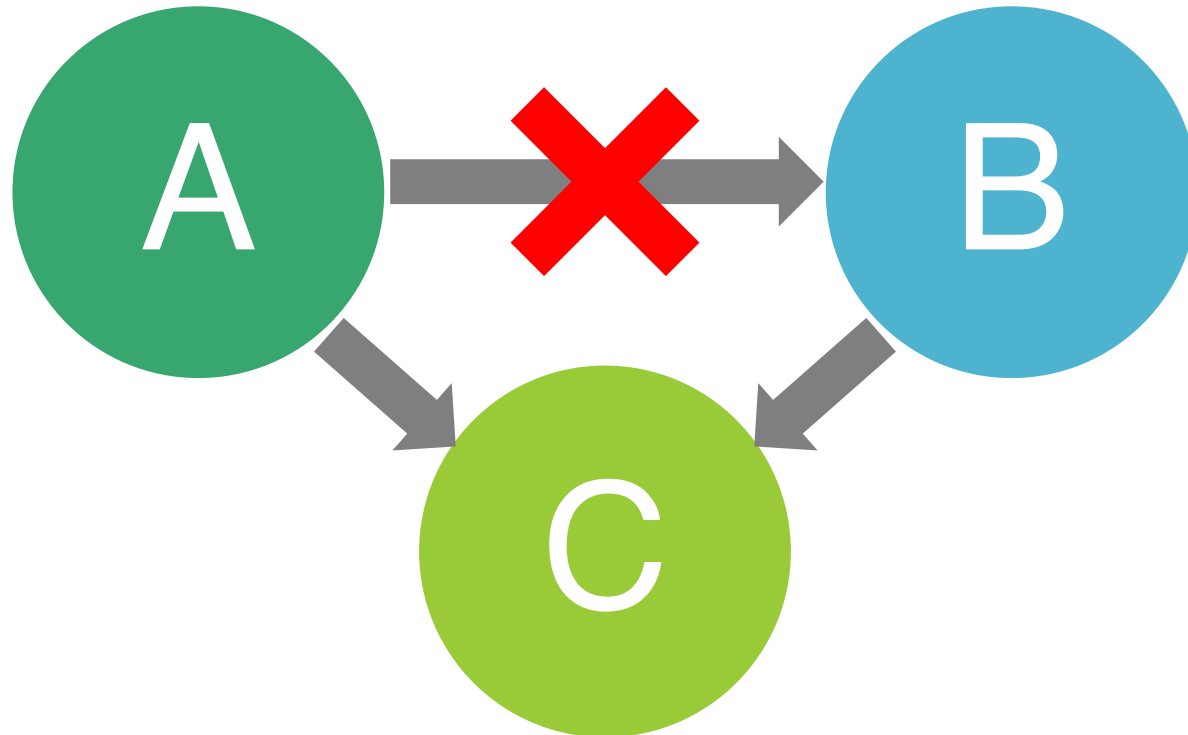
相関があっても因果関係がないケース②： 因果関係の方向が逆

- 例) 「事故の発生件数」と「注意喚起の看板の設置数」に正の相関があった



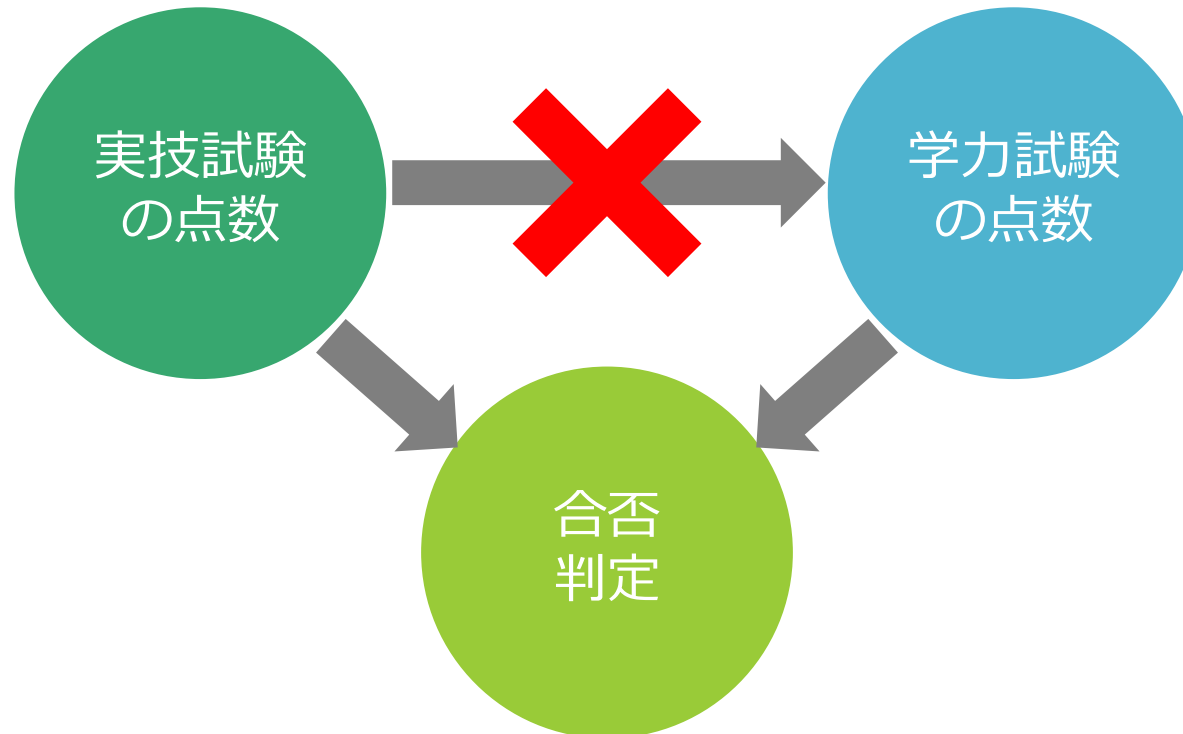
相関があっても因果関係がないケース③： 選択バイアス

- AとBの分析対象データが因果の合流点Cにおいて、選抜／層別／調整されてしまっている影響で、因果関係がないのに相関が生じてしまうケース



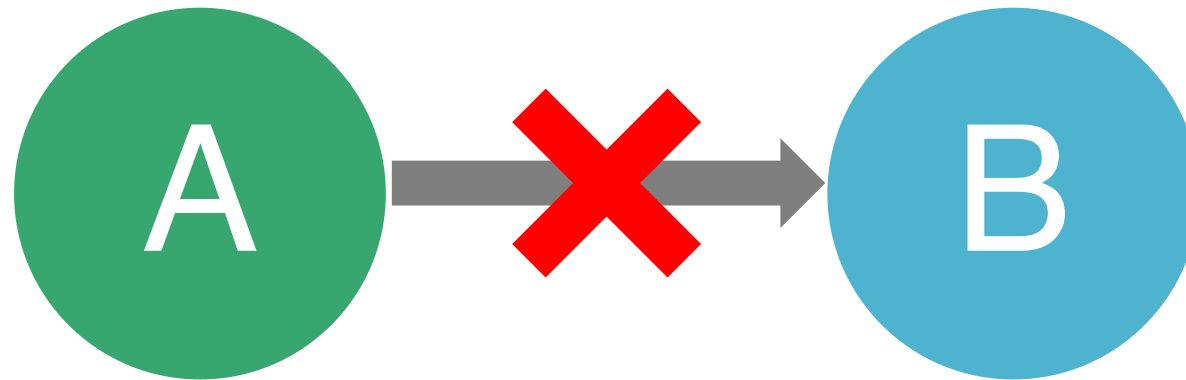
相関があっても因果関係がないケース③： 選択バイアス

- 例) 実技試験の点数と学力試験の点数の合計点で合否が決まる試験において、合格者の「実技試験の点数」と「学力試験の点数」に負の相関があった



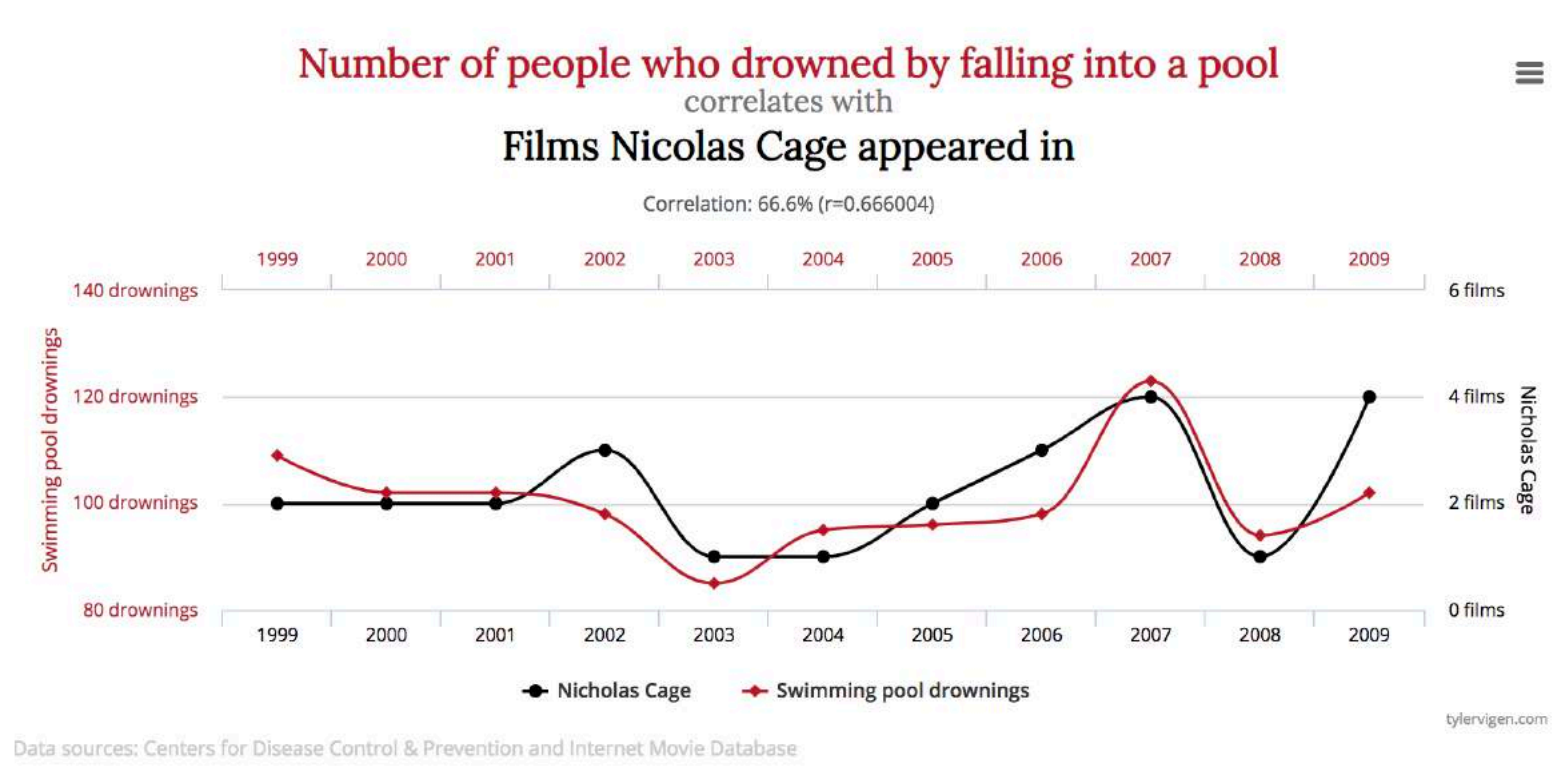
相関があっても因果関係がないケース④： 全くの偶然

- 全く何の関係もないAとBであっても、偶然相関が生じてしまうケース



相関があっても因果関係がないケース④： 全くの偶然

- 例) 「プールにおける溺死者数」と「ニコラス・ケイジの映画出演数」に正の相関があった



出典：<http://www.tylervigen.com/spurious-correlations>

データ活用におけるよくある勘違い

- ~~1. 定性データは役に立たない~~
→ 定量データと定性データは目的に応じて使い分ける
- ~~2. 相関があれば必ず因果関係がある~~
→ 相関があっても因果関係がない場合がある
3. エビデンスを求められた時は、
データを示せばOK!

データ活用におけるよくある勘違い

- ~~1. 定性データは役に立たない~~
→ 定量データと定性データは目的に応じて使い分ける
- ~~2. 相関があれば必ず因果関係がある~~
→ 相関があっても因果関係がない場合がある
3. エビデンスを求められた時は、
データを示せばOK!

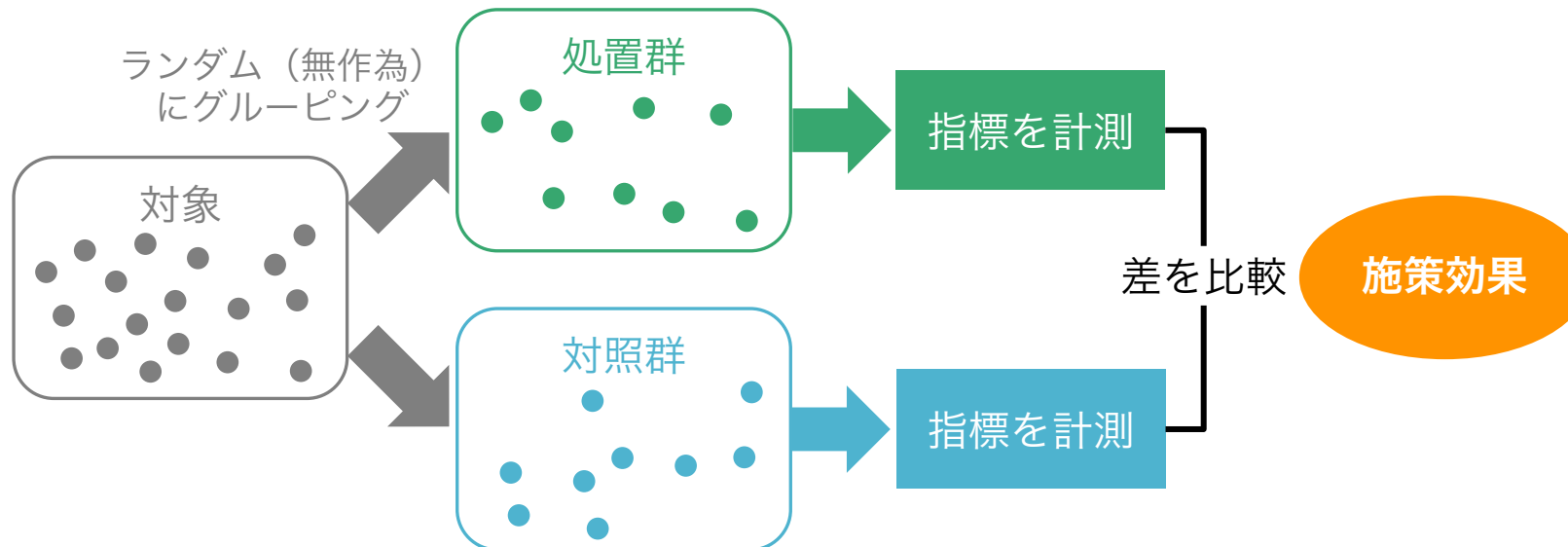
Before/Afterデータだけで 因果関係を推定するのは困難

- 実験室内なら、影響を及ぼす可能性のある要素を排除するように条件をコントロールできる
- しかし、実社会では特定の要素を排除することは困難
- Before/Afterのデータだけでは、施策によって課題が改善されたかは分からない（他の要因の可能性を排除できない）



ランダム化比較試験 (Randomized Controlled Trial: RCT)

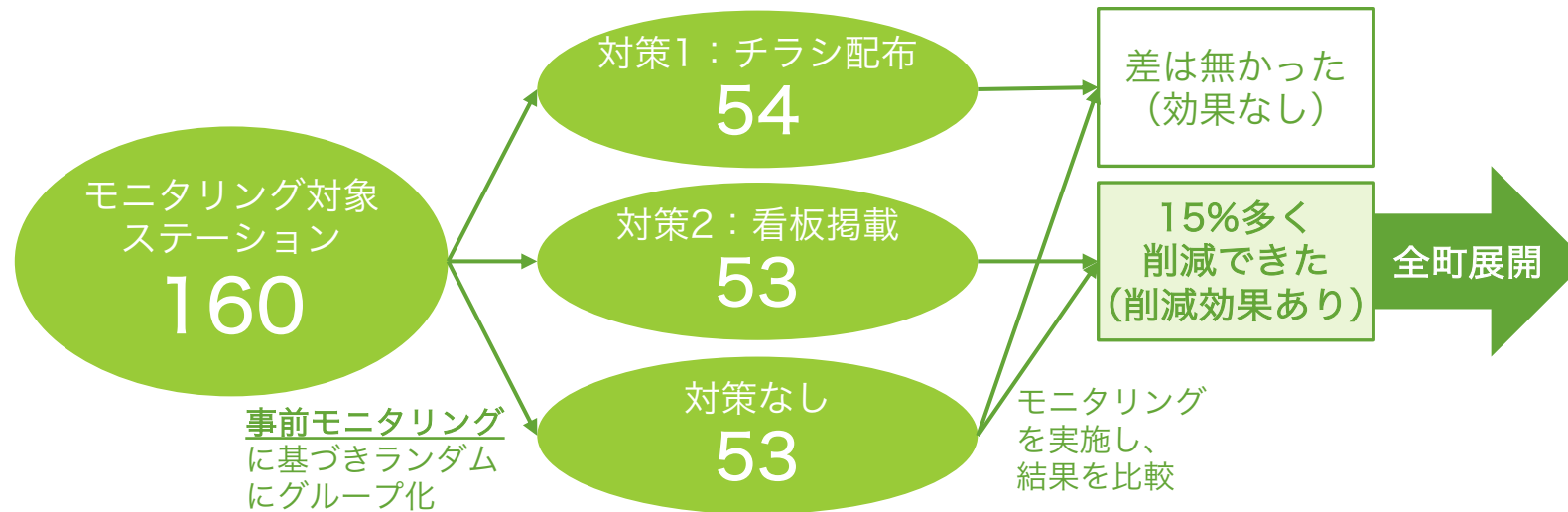
- 対象をランダムにグループ分けして指標を計測し比較
 1. 施策を実施するグループ (処置群)
 2. 施策を実施しないグループ (対照群)
- ランダム・グループピングによって他の要因による影響を排除



事例：神奈川県葉山町の 放置ごみ削減施策の効果検証

- 葉山町は2008年、日本で3番目に「ゼロ・ウェイスト政策」（ごみの徹底的な資源化・減量化）を導入
- 資源ごみの収集拠点である資源ステーションの管理は、ゼロ・ウェイストの達成に向けた重要施策
- 資源ステーションでの放置ごみが減らないことに町・住民が頭を悩ませていた

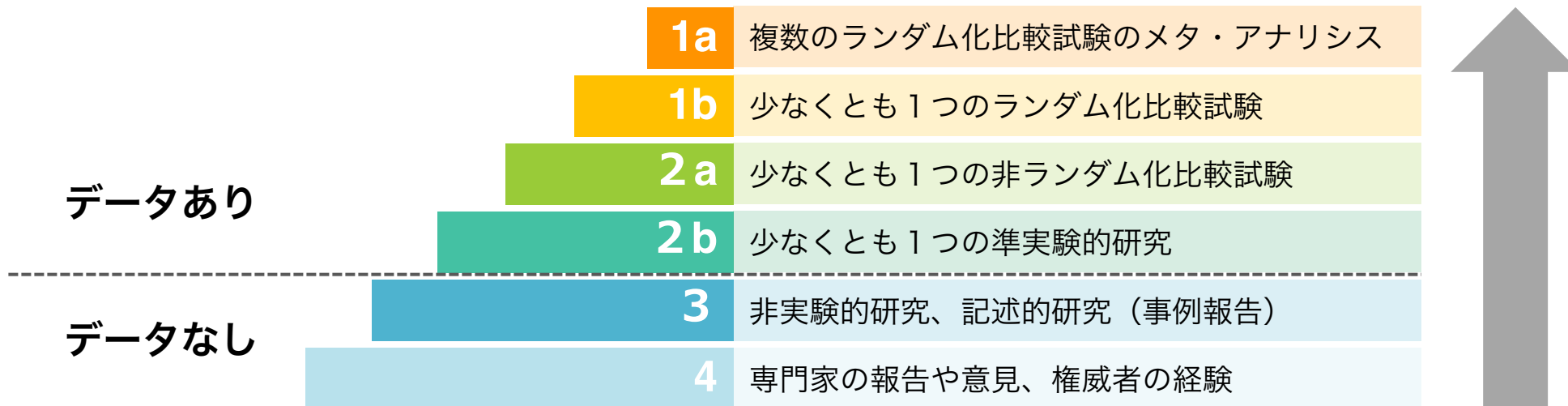
→ 施策実験を実施



出典：政策の効果をどう測定するか？：海外における「エビデンスに基づく政策」の最新動向
<https://www.rieti.go.jp/ip/events/bbl/16102501.html>

エビデンスの強さの階層

- 「エビデンスに基づく医療 (Evidence-Based Medicine)」に端を発したエビデンスの階層 (レベル) が社会科学にも応用されている
- レベルが上がるほどより正確に因果関係を推定できると見なされる



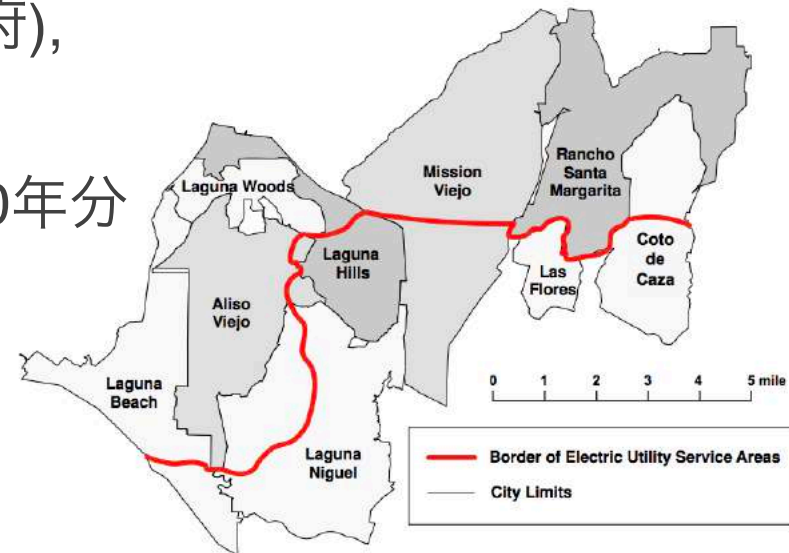
出典：正木朋也、津谷喜一郎（2006）「エビデンスに基づく医療（EMBの系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来）」『日本評価研究 第6巻第1号』pp.3-20

RCTが実施できない場合、 質の高いエビデンスを得ることは不可能？

- 答えは「No」
- ランダム比較化試験 (RCT) を行なったのと同様の状況のデータが入手できれば可能
- 既存の行政データが使える場合がある

例) カリフォルニア州における 電力消費量の分析

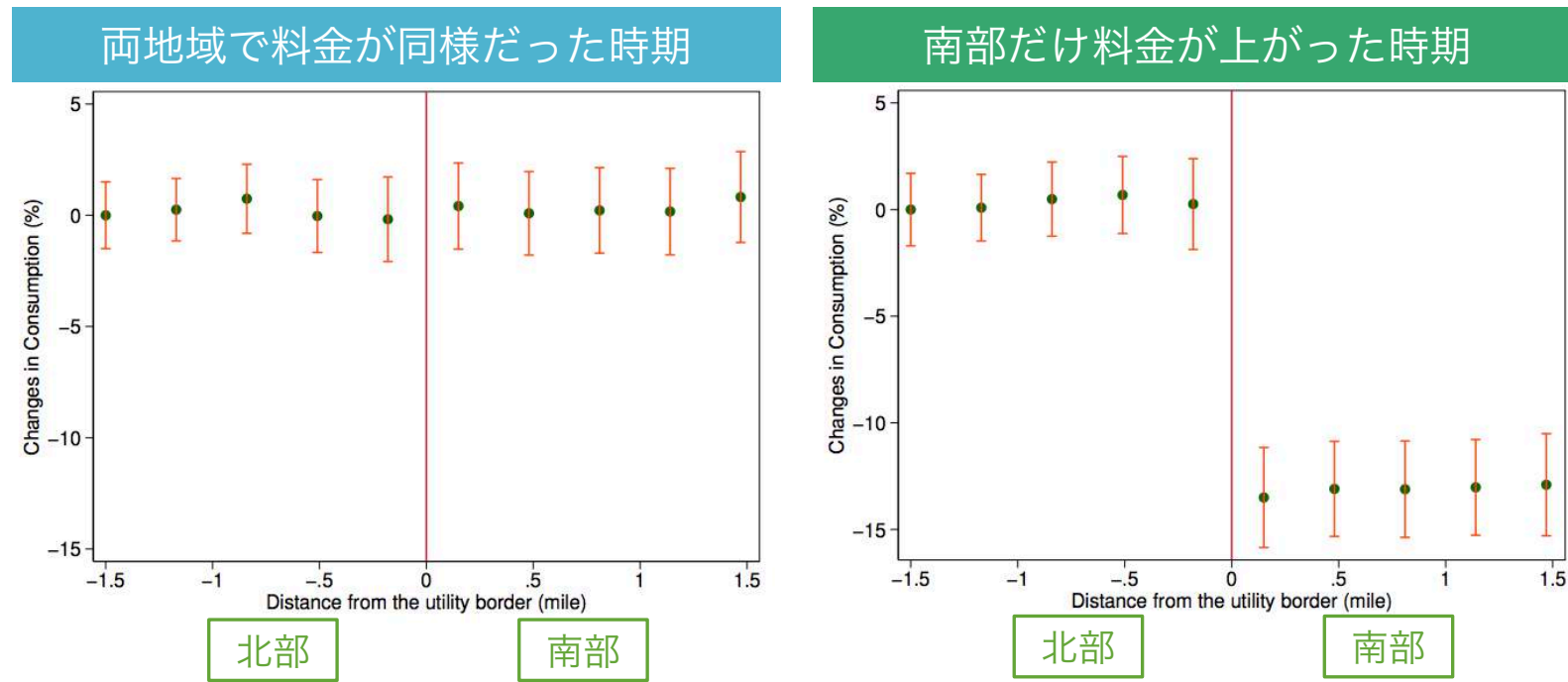
- 分析の目的：
「電力の料金を上げると、消費者がどのくらいの量を節電するのか知りたい」
- 分析実施者：シカゴ大学公共政策大学院 伊藤公一郎氏
- 協力機関：California Energy Commission (政府),
SCE, SDG&E (電力会社)
- 利用データ：世帯ごとの月間電力消費データ10年分
 - カリフォルニア州のある都市では南北で電力会社が異なる
 - 両方の電力会社の電力料金が同様に変化していた時期と、南部だけ電力料金が上がった時期がある



出典：政策の効果をどう測定するか？：海外における「エビデンスに基づく政策」の最新動向
<https://www.rieti.go.jp/jp/events/bbl/16102501.html>

例) カリフォルニア州における 電力消費量の分析

両方の電力会社の電力料金が同様に変化していた時期と比較すると、南部で電力料金が上がった
時期の電力消費量は12-13%減少していた
→電力料金が上がると消費者は節電する



データ出典 : Koichiro Ito, American Economic Review 2014, 104(2): 537-563
<http://dx.doi.org/10.1257/aer.104.2.537>

データ活用における よくある勘違い

- ~~1. 定性データは役に立たない~~
→ 定量データと定性データは目的に応じて使い分ける
- ~~2. 相関があれば必ず因果関係がある~~
→ 相関があっても因果関係がない場合がある
- ~~3. エビデンスを求められた時は、
データを示せばOK!~~
→ エビデンスにも質のレベルが存在するため、
ただデータを示すだけでは根拠として弱い場合がある

5. 行政プロセスにデータ 分析を取り入れる上で おすすめの文献

データを活用した政策立案プロセスを学びたい方へ データ活用で地域のミライを変える！ 課題解決の7Step

NEW

おすすめ 試し読み



データ活用で地域のミライを変える！
課題解決の7 Step

紙：¥ 2,310
 電子：¥ 2,310

数量
1

 カートに入れる

 お気に入りに追加

クレジットカード払い・コンビニ支払い・請求書払い対応

編著者名 : 一般社団法人 コード・フォー・ジャパン / 編著
判型 : A5・196ページ
商品形態 : 単行本
雑誌コード :
図書コード : 5108579-00-000
8172327-00-000
ISBNコード : 978-4-324-10764-5
発行年月 : 2019/12
販売価格 : 2,310円 (税込み)

試し読み >

- 50以上の自治体を受講したデータ活用研修
“データアカデミー”のエッセンスを1冊に集約
- 2019年12月に発売されたばかりの新刊

<https://shop.gyosei.jp/products/detail/10196>

人に「伝える」ためのデータ可視化の基礎を押さえない方へ グラフをつくる前に読む本



グラフをつくる前に読む本 一瞬で伝わる表現はどのように生まれたのか 単行本（ソフトカバー） - 2017/9/23

松本 健太郎 (著)

★★★★☆ (31件のカスタマーレビュー)

その他 (2) の形式およびエディションを表示する

Kindle版
¥ 1,642

単行本（ソフトカバー）
¥ 1,728 ✓prime

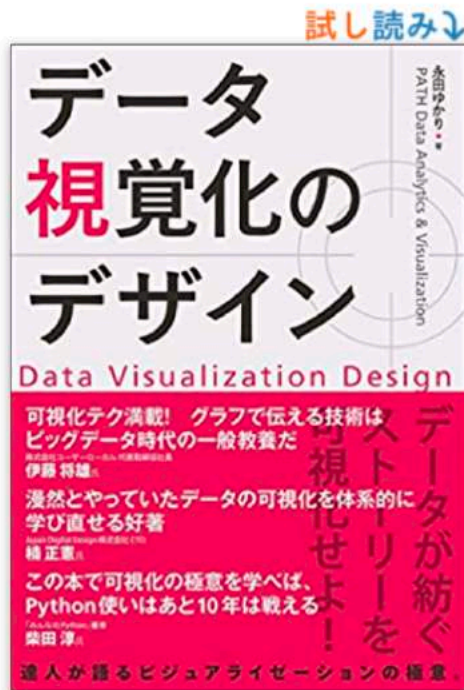
今すぐお読みいただけます: **無料アプリ**

¥ 667 より 26 中古品の出品
¥ 1,728 より 4 新品



https://www.amazon.co.jp/dp/4774192198/ref=cm_sw_r_tw_dp_U_x_JgZnDbPN2PAXN

さらに「わかりやすく伝える」ためのノウハウを学びたい方へ データ視覚化のデザイン



データ視覚化のデザイン (日本語) 単行本 - 2020/6/19

永田 ゆかり (著)

★★★★☆ 46個の評価

その他 の形式およびエディションを表示する

Kindle版 (電子書籍)
¥1,980
獲得ポイント: 20pt

単行本
¥2,200
獲得ポイント: 22pt ✓prime

今すぐお読みいただけます: **無料アプリ**

¥3,015 より 3 中古品
¥2,200 より 8 新品

この商品の特別キャンペーン SUUMO住宅情報誌と対象の本の同時購入で合計金額から最大370円OFF。 1件

10月13の火曜日、8AM-12PMの間にお届けします。 購入手続き画面で都合がいい時間帯を選択してください [詳細](#)

最速の配達日: 10月13の火曜日、8AM-12PMの間にお届けします。 購入手続き画面でご都合のよい時間帯を選択してください。

達人が教えるビジュアライゼーションの極意!!

https://www.amazon.co.jp/dp/4815604053/ref=cm_sw_em_r_mt_dp_mx0GFb8P54REP

データ活用型組織を作るためのヒントを得たい方へ ワーカーマン式「しない経営」



ワーカーマン式「しない経営」—— 4000億円の空白市場を切り拓いた秘密 (日本語) 単行本 (ソフトカバー) – 2020/10/21
土屋 哲雄 (著)

[その他](#) の形式およびエディションを表示する

Kindle版 (電子書籍)
¥1,584
獲得ポイント: 16pt

単行本 (ソフトカバー)
¥1,760
獲得ポイント: 18pt

今すぐお読みいただけます: **無料アプリ**

¥1,760 より 1 新品

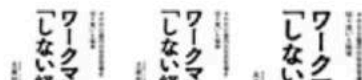
「予約商品の価格保証」対象商品。 [詳細](#)

この商品の特別キャンペーン SUUMO住宅情報誌と対象の本の同時購入で合計金額から最大370円OFF。 1 件

☆本書で対談した早稲田大学大学院・ビジネススクール 入山章栄教授 大絶賛!

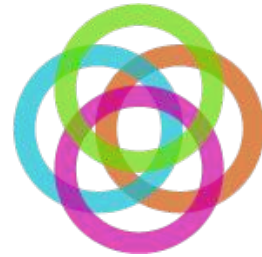
☆急成長ワーカーマンの仕掛け人、待望の処女作!

☆残業・ノルマ・期限一切なし!



https://www.amazon.co.jp/gp/product/4478111456/ref=ppx_yo_dt_b_asin_title_o00_s00

Link Data Now!!



LinkData.org

本資料に関するご質問・ご意見や、
データ活用研修・ワークショップ開催のご相談などは
こちらへお寄せください。

一般社団法人リンクデータ
代表理事 下山 紗代子
Email: support@linkdata.org