

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

2018年度 政策評価に関する統一研修

一般社団法人リンクデータ 代表理事 下山 紗代子

目次

1. 行政データ分析を行う上で覚えておきたい考え方 …P.4
2. 行政データ活用に立ちはだかる壁とその乗り越え方 …P.52
3. データ分析の流れと分析手法の選び方 …P.64
4. 公共データの活用がもたらすインパクト …P.81

講師：下山 紗代子（しもやま さよこ）



「データフル活用社会」の実現を目指し、国内最大級のオープンデータ活用支援プラットフォーム「LinkData.org」の運営や、日本初のデータ活用コンテスト「Liked Open Data チャレンジ Japan」の運営、その他データ活用イベントやデータ分析ワークショップ等の実施を通し、データ活用人材の育成に取り組んできた。元々は生物学に興味があり、大学時代は沖縄で熱帯性魚類の研究や深海調査を行っていた。卒業後はバイオインフォマティクス系ベンチャーのシステムエンジニアを経て、理化学研究所に入所。生命科学統合データベース構築事業において、研究者のためのデータ公開支援システムを開発したことをきっかけに、異分野のデータを組み合わせることの面白さに目覚め、オープンデータに興味を持つ。

その後研究開発したデータ公開技術を元に独立し、2014年に一般社団法人リンクデータを設立。現在は同社団代表理事を務める他、Liked Open Data チャレンジ Japan 実行委員長、総合地球環境学研究所共同研究員、Code for Japan アドバイザー、ミーカンパニー株式会社 データスチュワード、岩崎学園情報科学専門学校非常勤講師、インフォ・ラウンジ株式会社 取締役、総務省地域情報化アドバイザーを務めている。

趣味は所有するアメリカンバイクでのツーリング。



LinkData.org

1. 行政データ分析を行う上で 覚えておきたい考え方

データ分析に必要なのは 「サイエンティスト思考」

2018年 ノーベル医学生理学賞 本庶佑氏

「教科書を信じるな」



「何でも疑え」というわけではない

与えられた情報を全て鵜呑みにするのではなく、

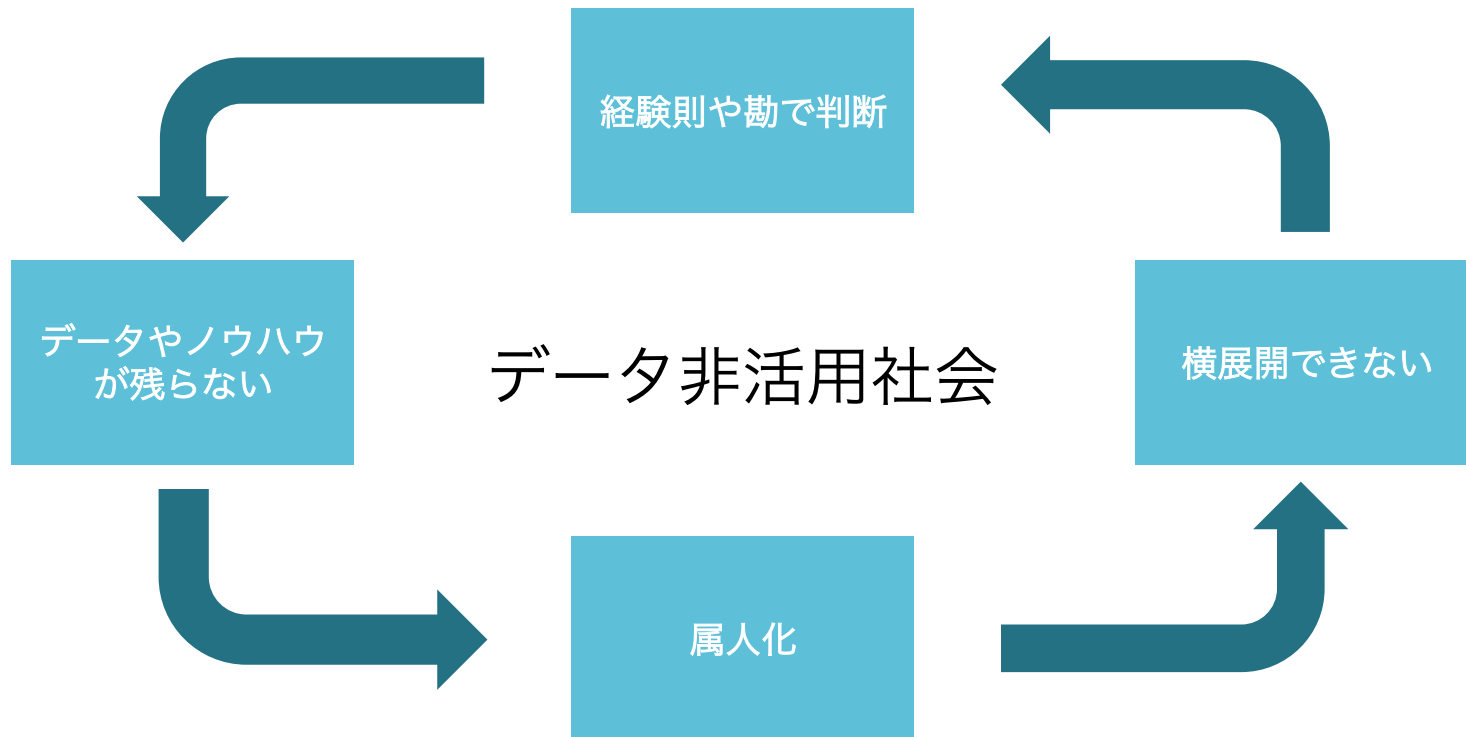
「本当にそうなのだろうか？」

「なぜそうなのだろうか？」

と自分で考える姿勢 = **サイエンティスト思考** が重要

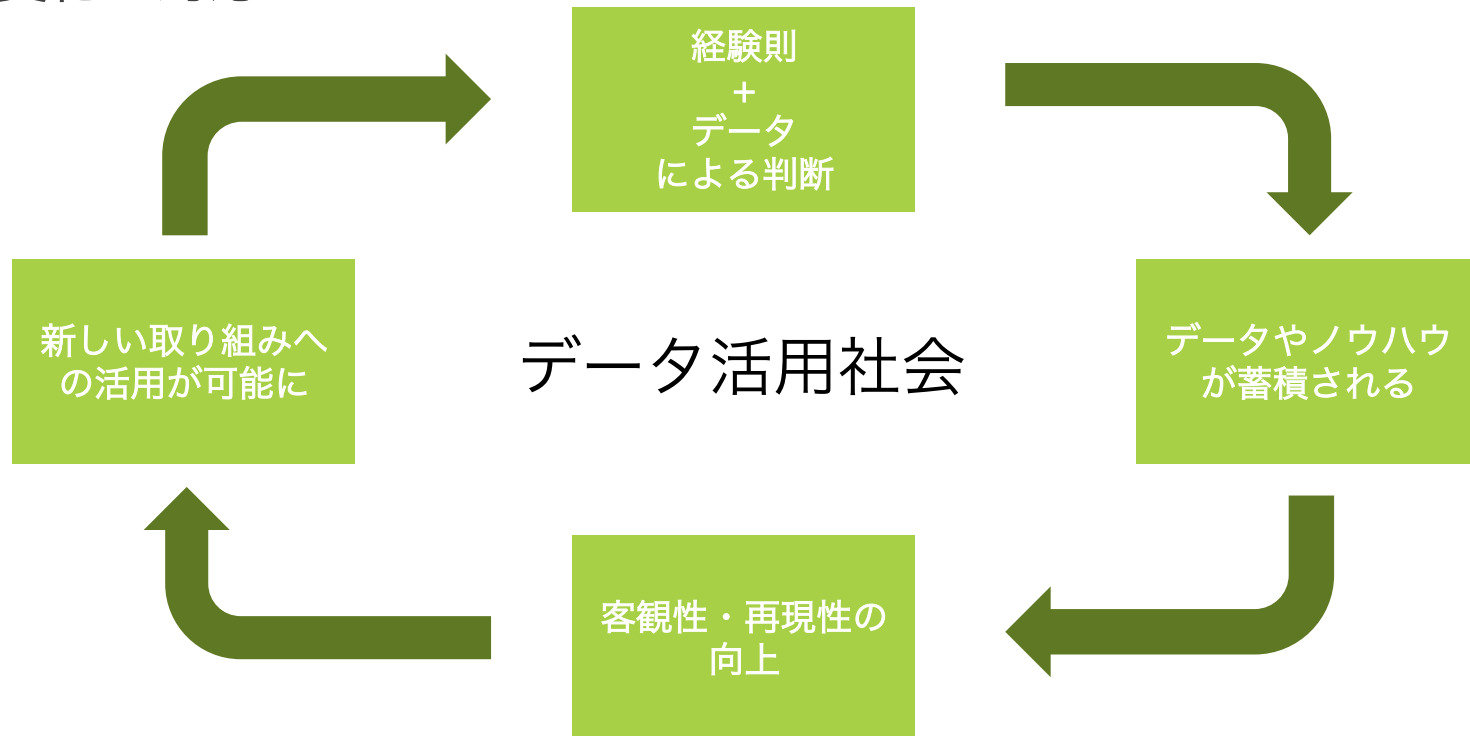
データが活用されない 負のループ

現状維持では環境の変化に対応できず、縮小する一方



データ利活用によって 正のループを回す

データの蓄積 + 新たに得られたデータを取り入れて
変化に対応できる





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

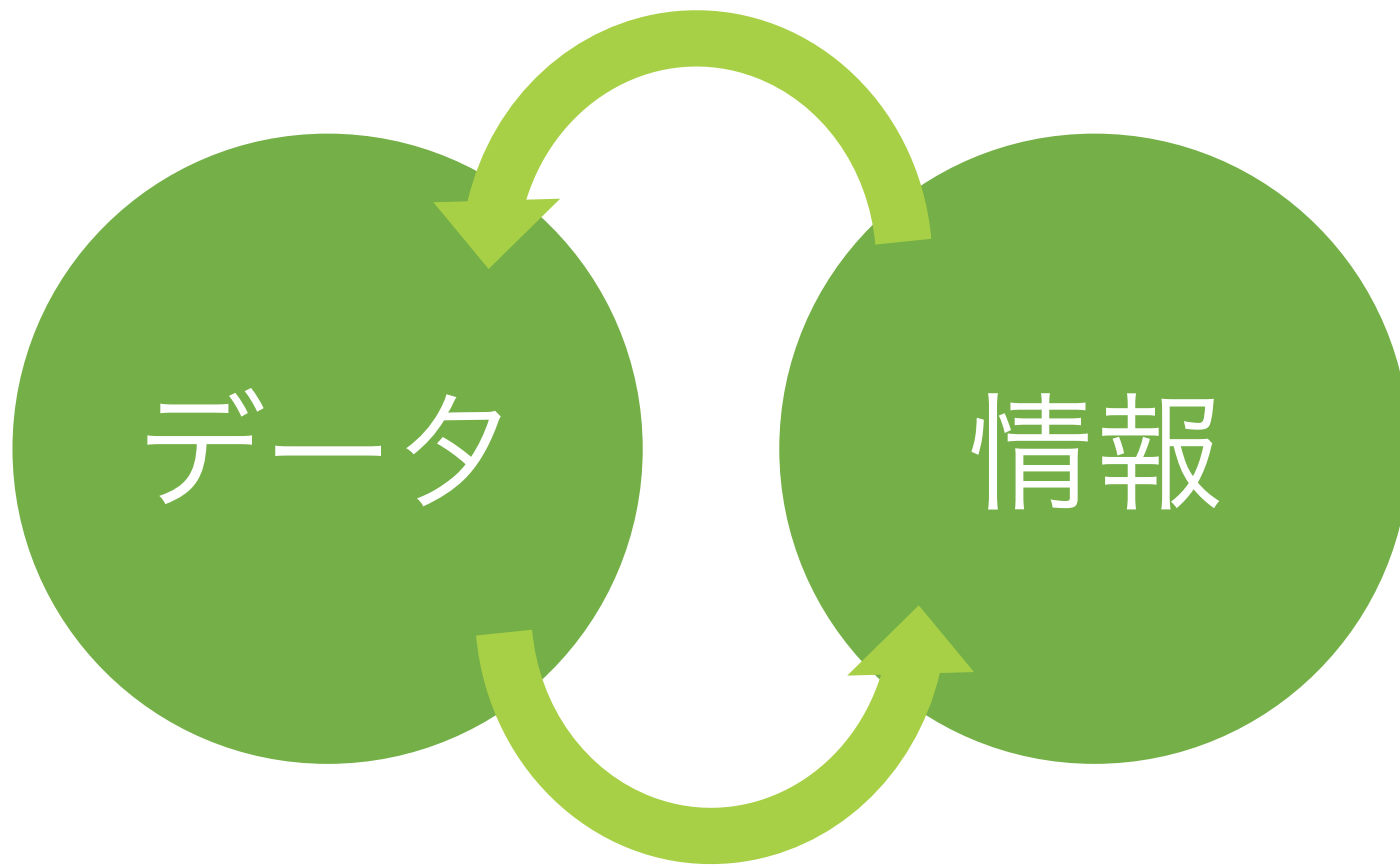
"Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning.

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

2017

4087

数字

2017年の
宇都宮市における
ぎょうざの年間の支出金額
(1世帯当たり)は
4,087円です。

情報

市区町村名	調査年	ぎょうざの年間支出金額 (1世帯当たり)
宇都宮市	2017	4087

データ

情報⇔データ 双方向に変換可能

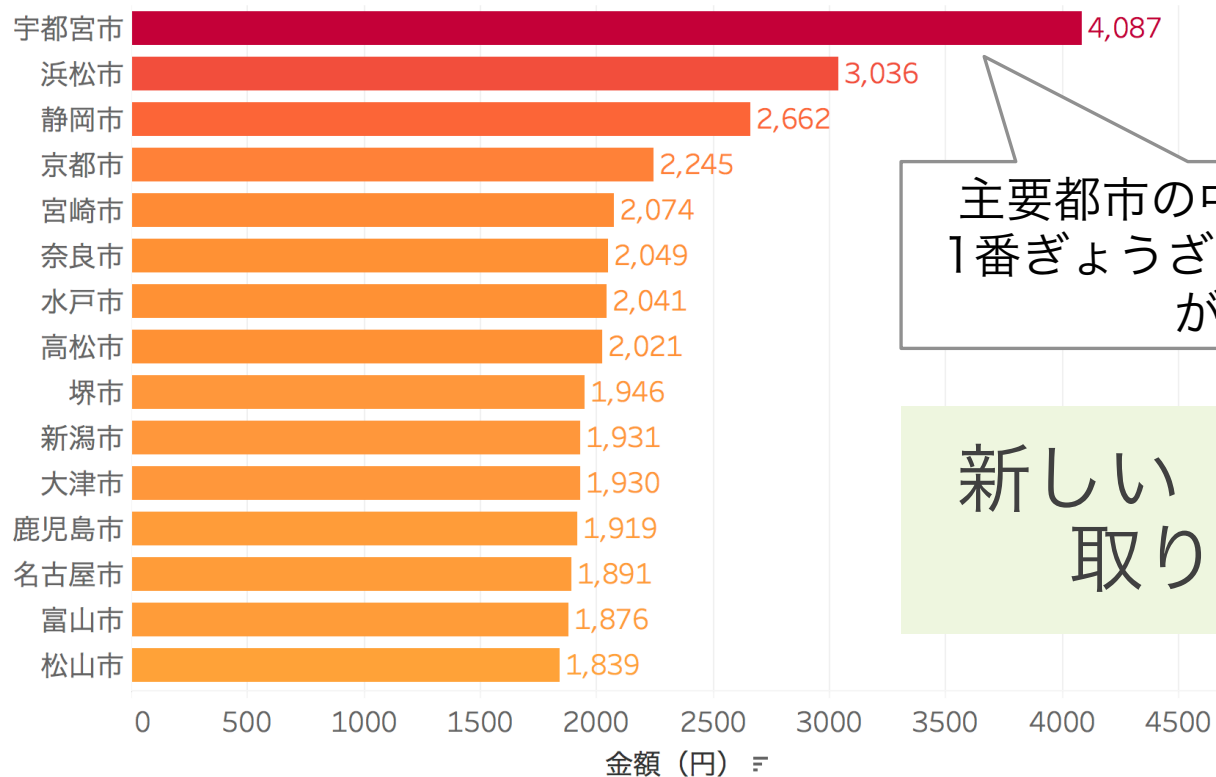
2017年の
宇都宮市における
ぎょうざの年間の支出金額
(1世帯あたり) は
4,087円です。



市区町村名	調査年	ぎょうざの年間 支出金額 (1世 帯あたり)
宇都宮市	2017	4087

データ化することで 新しい「情報」が取り出せる

主要都市におけるぎょうざの年間支出金額（一世帯あたり）



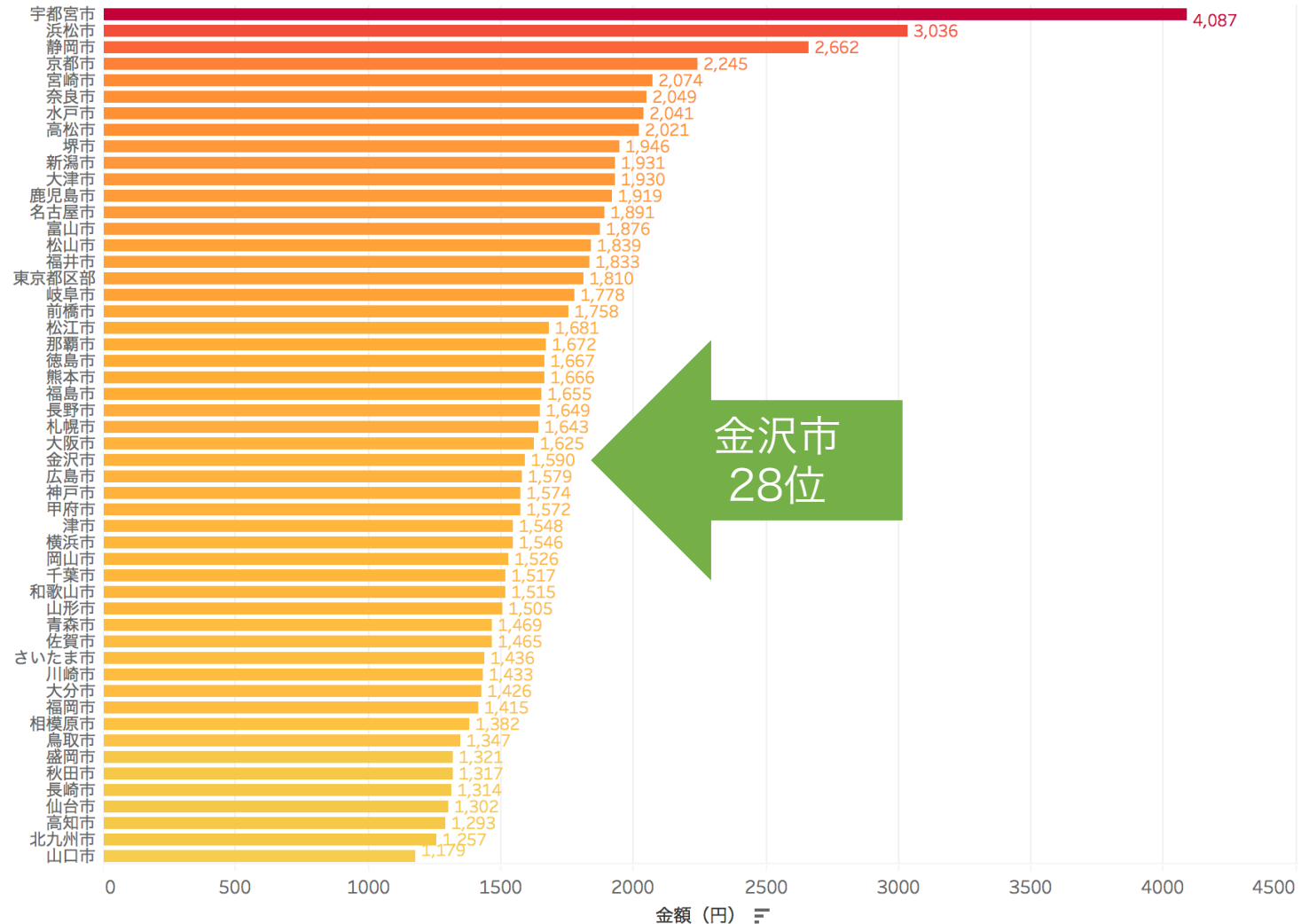
主要都市の中で宇都宮市が
1番ぎょうざの年間支出金額
が多い

新しい「**情報**」が
取り出せる

データ出典元：総務省家計調査（家計収支編）調査結果 > 総世帯 > 2017年 >（品目分類）1世帯当たり品目別支出金額 > 11.都市階級・地方・都道府県庁所在市別

(おまけ) フルバージョン

主要都市におけるぎょうざの年間支出金額 (1世帯当たり)



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

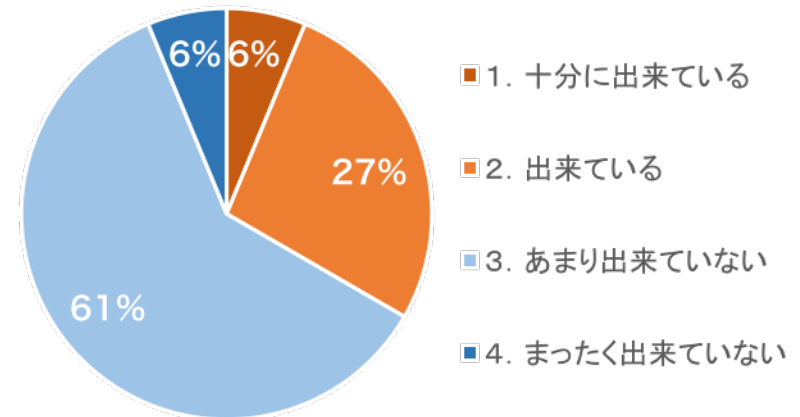
あなたは、日々の業務の中で、データを活用出来ていると思いますか？

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

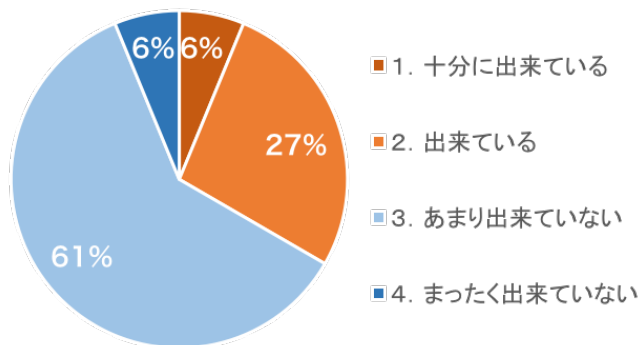


定量データの使いどころは「俯瞰」

- 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

研修実施前

Q: 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？

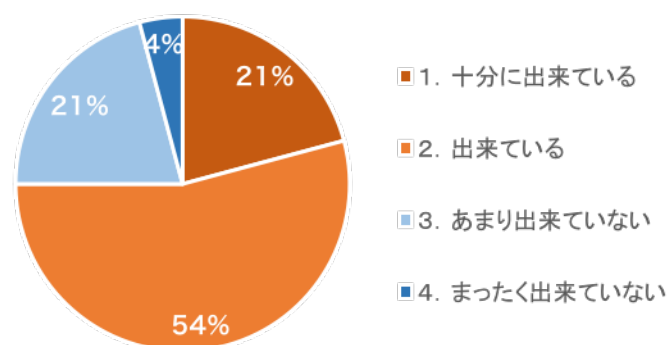


全受講者の **33%** が日常の業務でデータを活用できている

研修の
実施

研修実施1年後

Q: 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？



全受講者の **75%** が日常の業務でデータを活用できている

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）
普通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が多いため、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
...

分類・
整理

ピックアップ

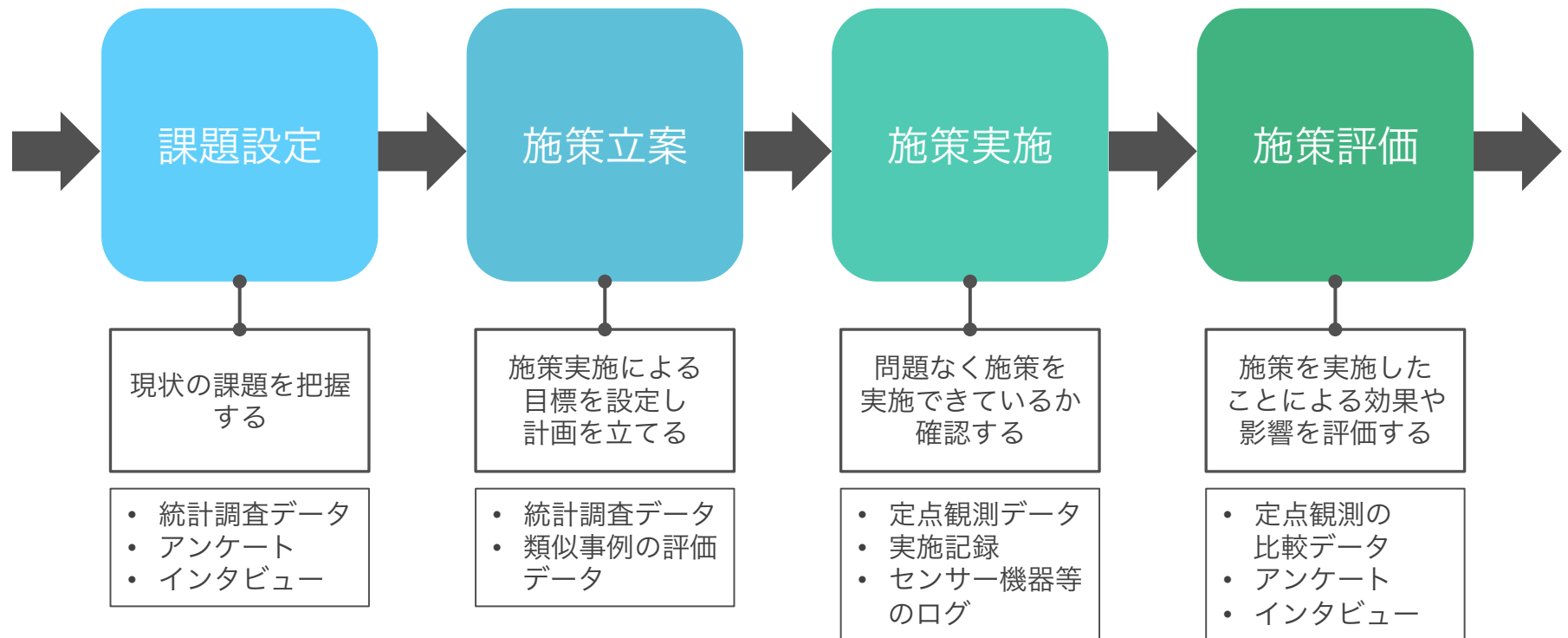
データを自由に使える環境が整っていない

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

定量データ・定性データは うまく組み合わせる必要がある



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うまく使い分ける
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！

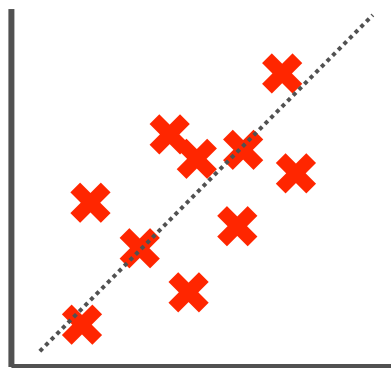
データ活用におけるよくある勘違い

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うまく使い分ける
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

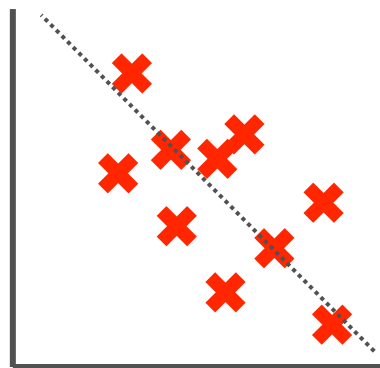
二つの事象が密接にかかわり合い、一方が変化すれば他方も変化するような関係。

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

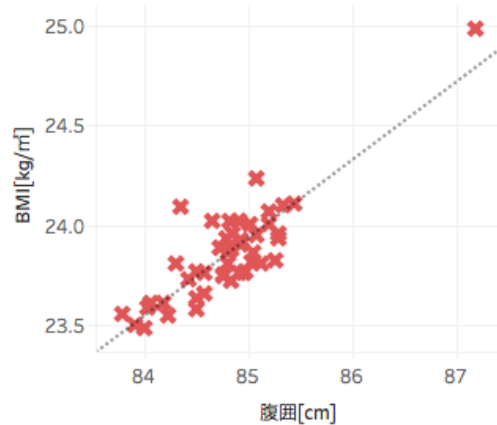
負の相関



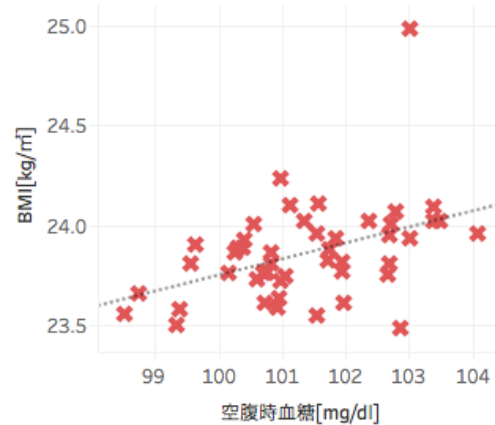
一方の値が大きくなると、
もう一方の値は小さくなる

相関の強さの比較

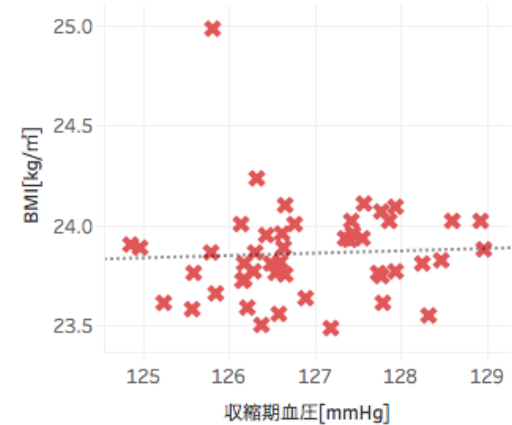
非常に強い正の相関 ($r=0.86$)



強い正の相関 ($r=0.43$)



ほぼ相関がない ($r=0.05$)



<https://public.tableau.com/views/22956/sheet4>

相関係数 (r) の値の範囲

解釈

$0.7 \leq r \leq 1.0$

非常に強い正の相関がある

$0.4 \leq r < 0.7$

強い正の相関がある

$0.2 \leq r < 0.4$

やや正の相関がある

$-0.2 < r < 0.2$

ほぼ相関がない

$-0.4 < r \leq -0.2$

やや負の相関がある

$-0.7 < r \leq -0.4$

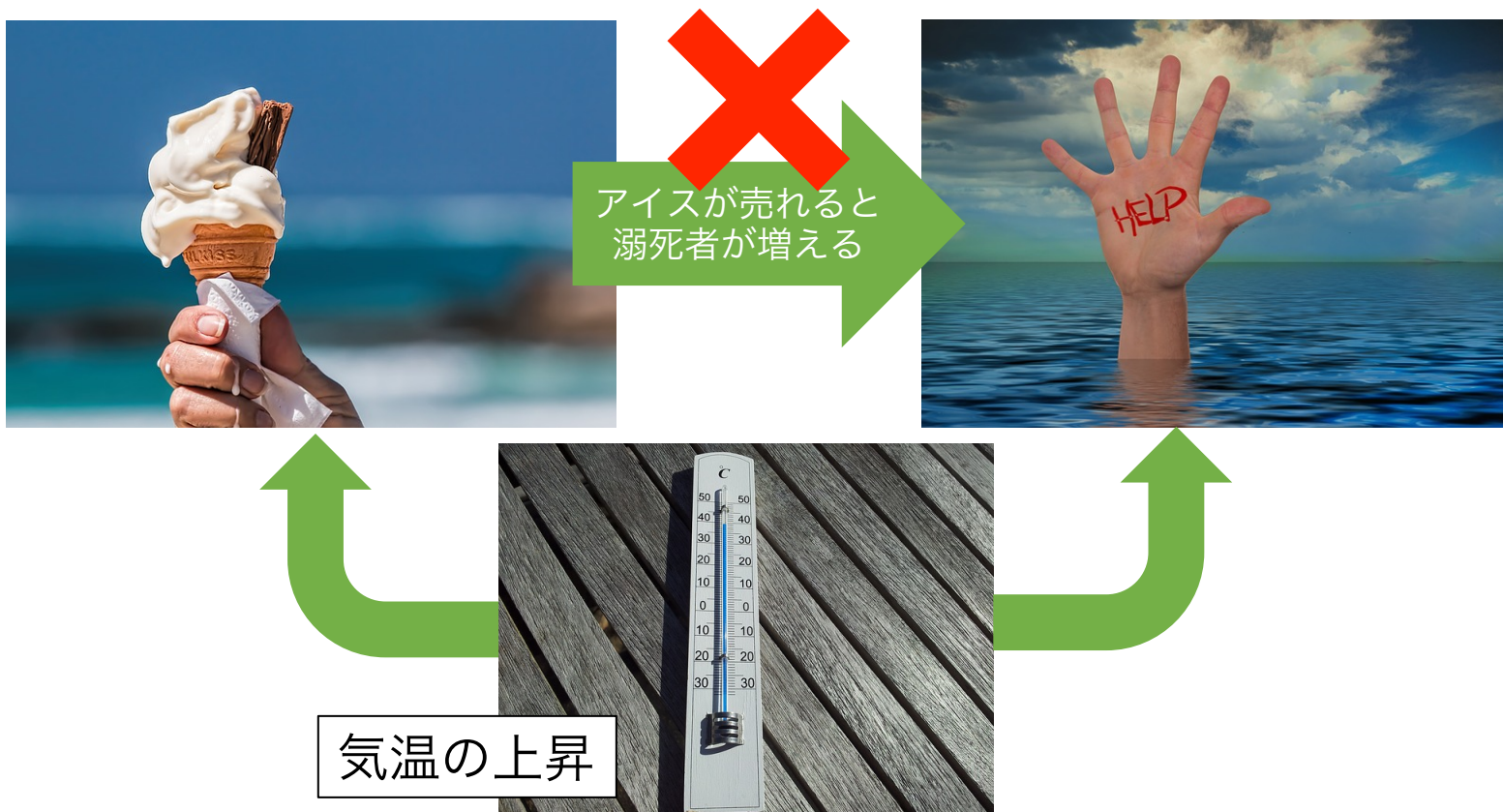
強い負の相関がある

$-1.0 \leq r \leq -0.7$

非常に強い負の相関がある

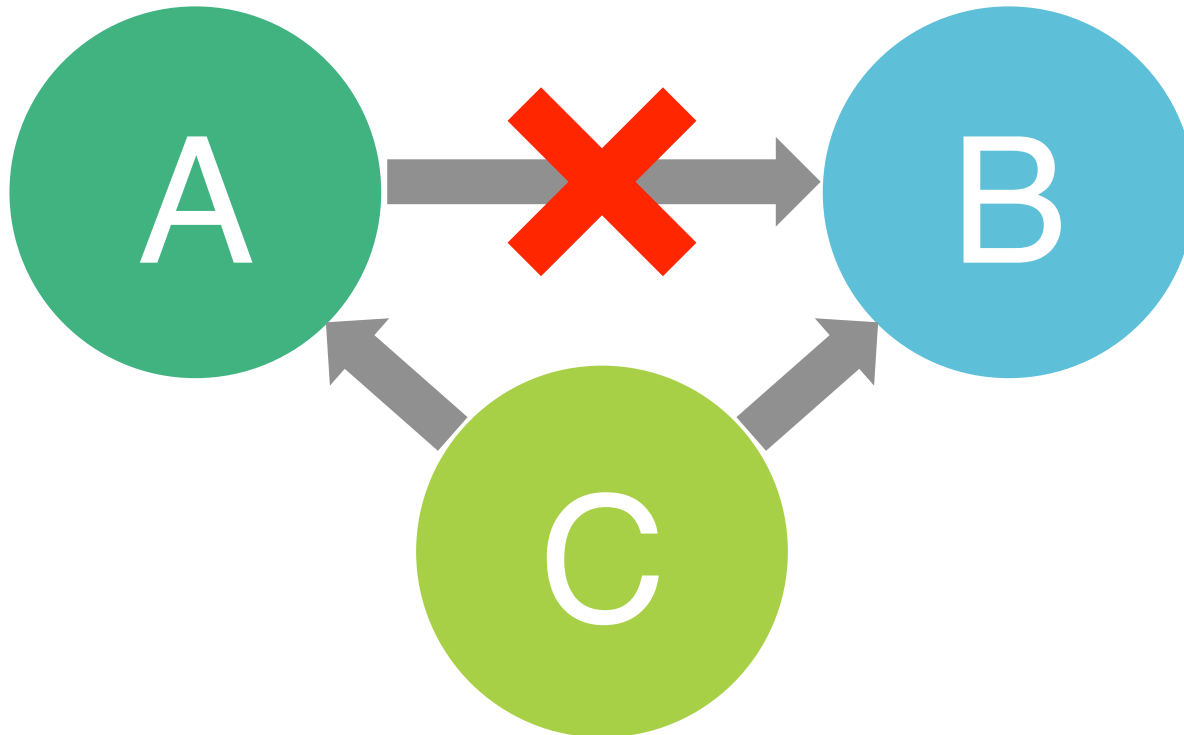
相関≠因果関係

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



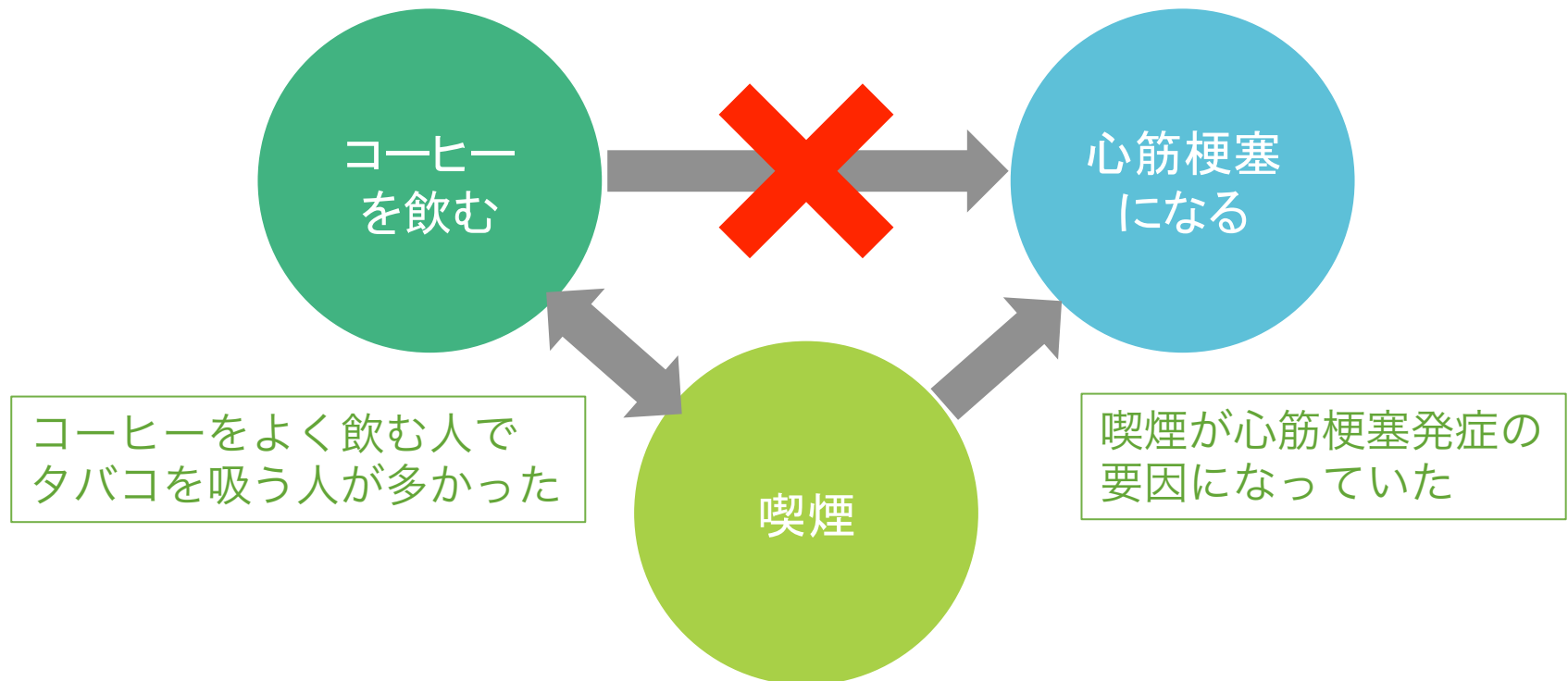
相関があっても因果関係がないケース①： 共通の要因がある（交絡）

- 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



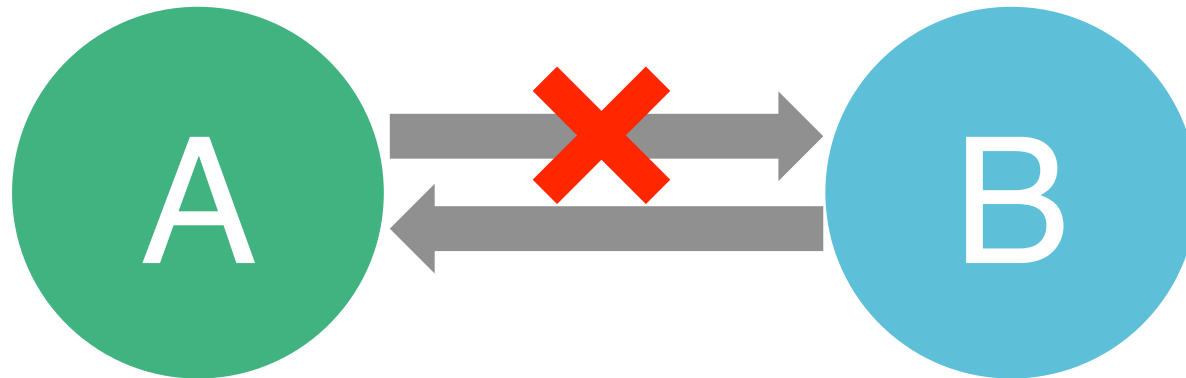
相関があっても因果関係がないケース①： 共通の要因がある（交絡）

- 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



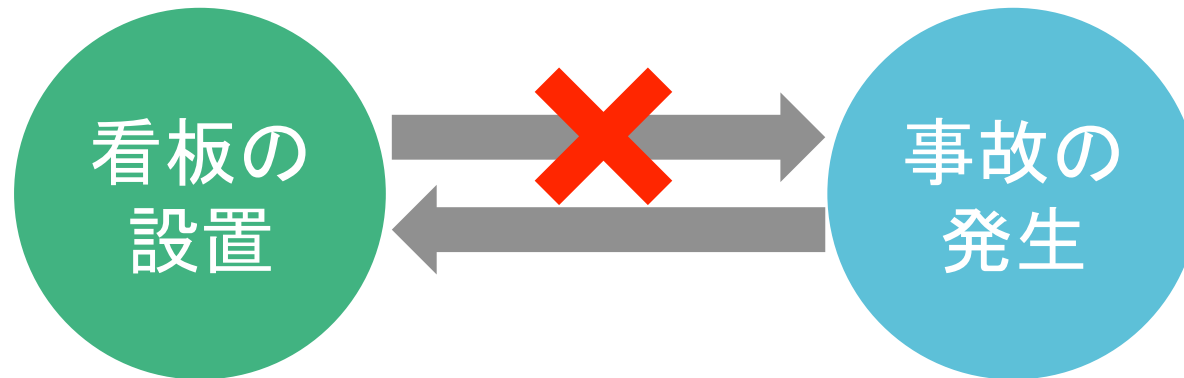
相関があっても因果関係がないケース②： 因果関係の方向が逆

- 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



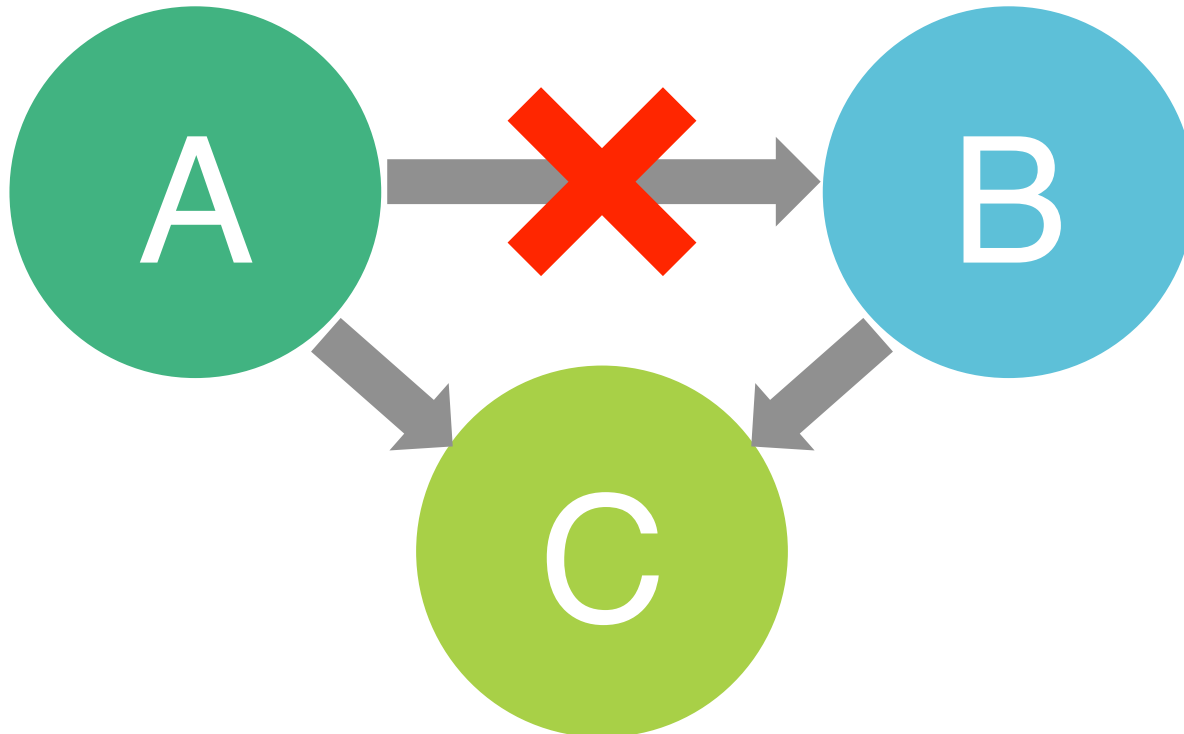
相関があっても因果関係がないケース②： 因果関係の方向が逆

- 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



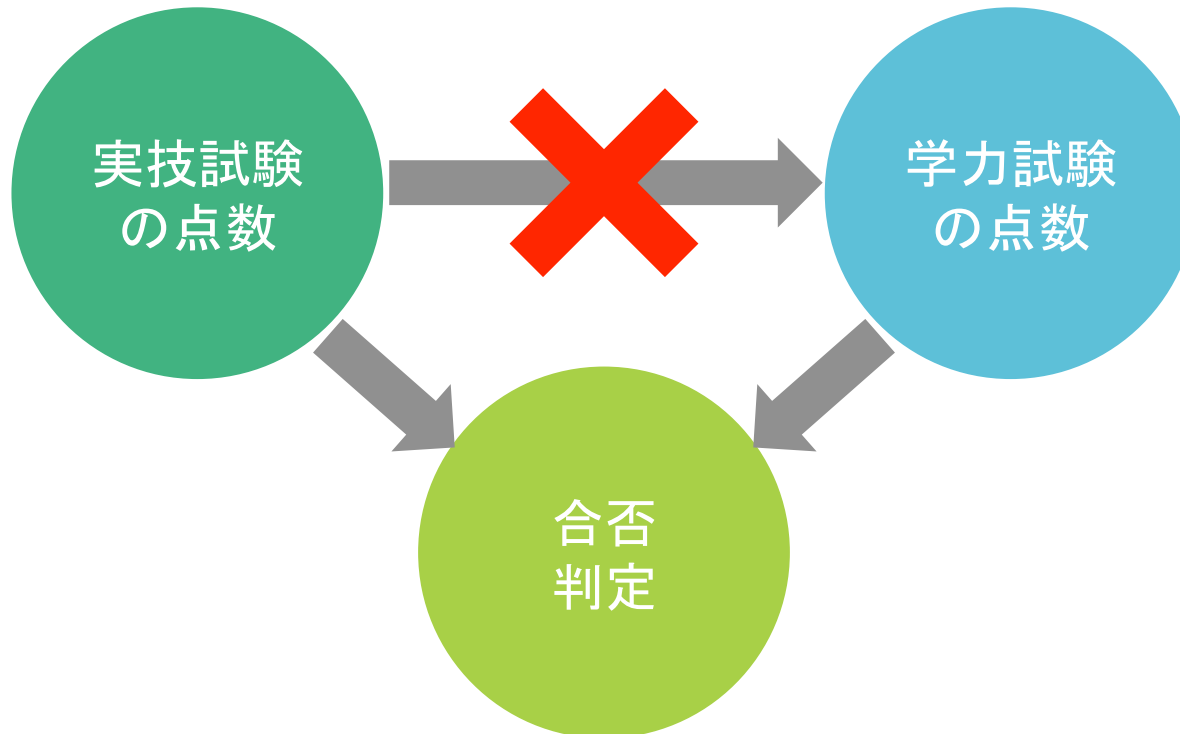
相関があっても因果関係がないケース③： 選択バイアス

- 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



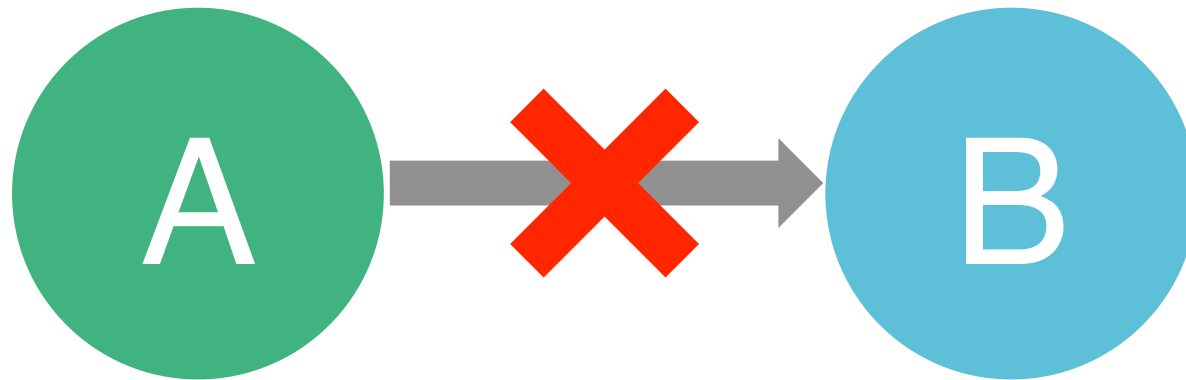
相関があっても因果関係がないケース③： 選択バイアス

- 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



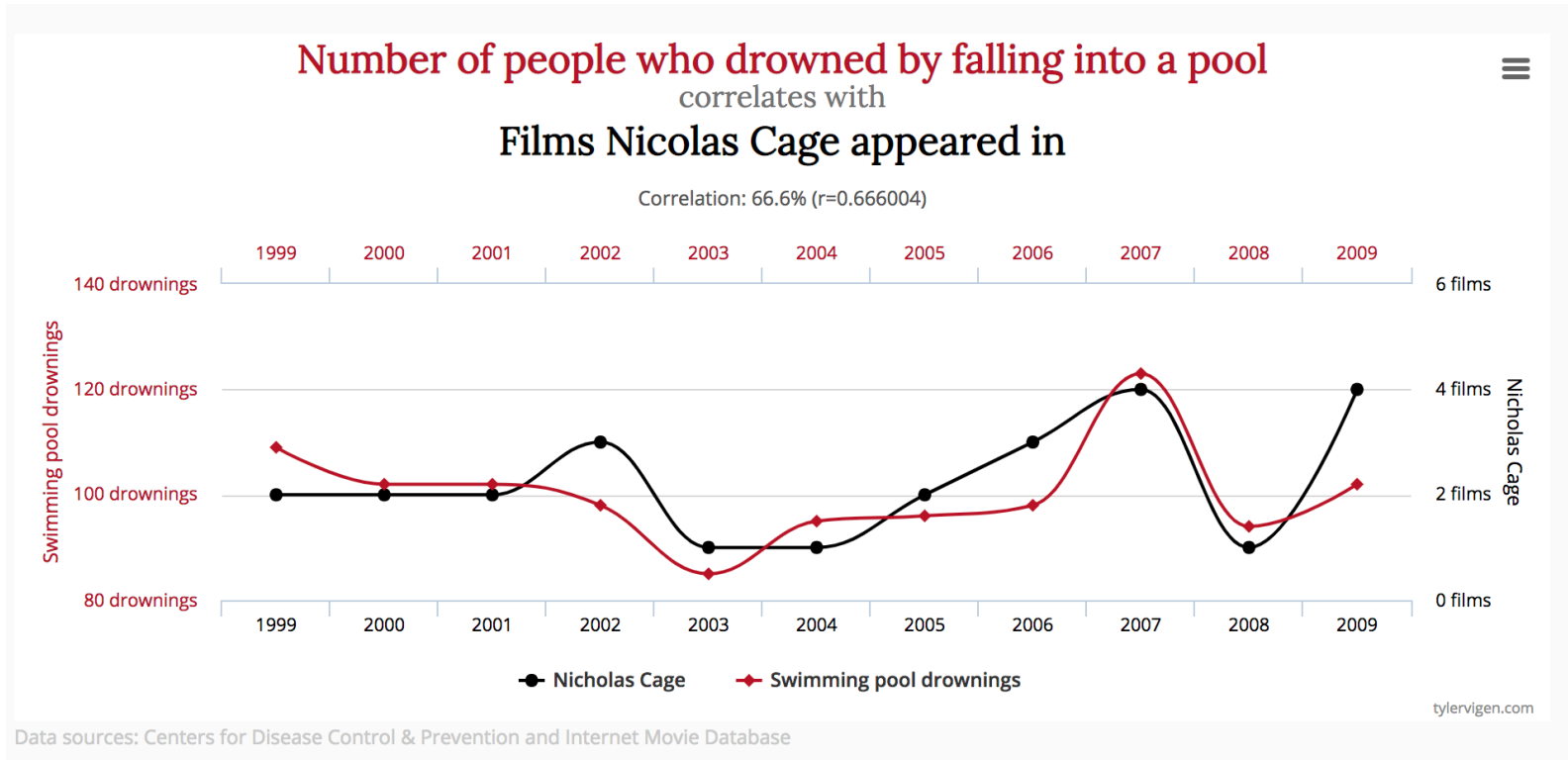
相関があっても因果関係がないケース④： 全くの偶然

- 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

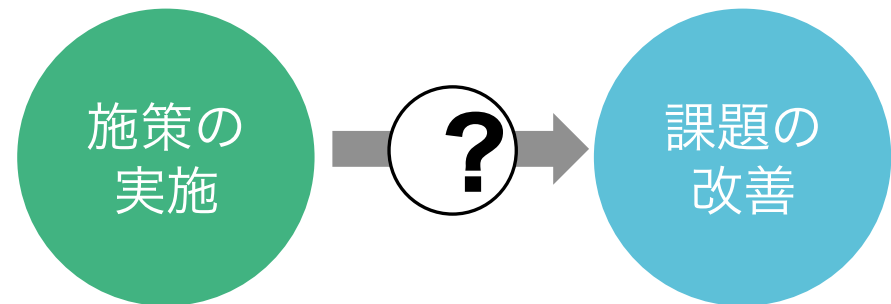
1. ~~定性データは役に立たない~~
→ 定量データと定性データはうまく使い分ける
2. ~~相関があれば必ず因果関係がある~~
→ 相関があっても因果関係がない場合がある
3. エビデンスを求められた時は、
データを示せばOK！

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うまく使い分ける
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！

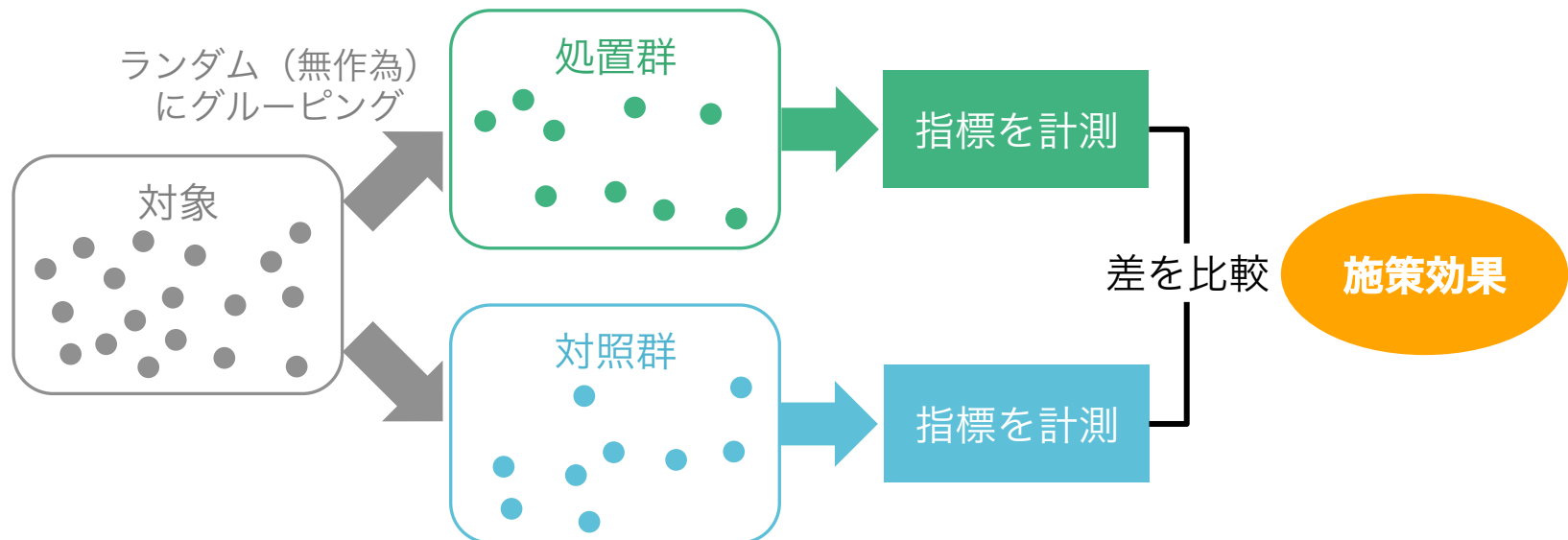
Before/Afterデータだけで因果関係を特定するのは困難

- 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
- 大量のデータや高度な統計モデルがあったら解決可能？
→ 難しいと言われている
- したがって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）



ランダム比較化試験 (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

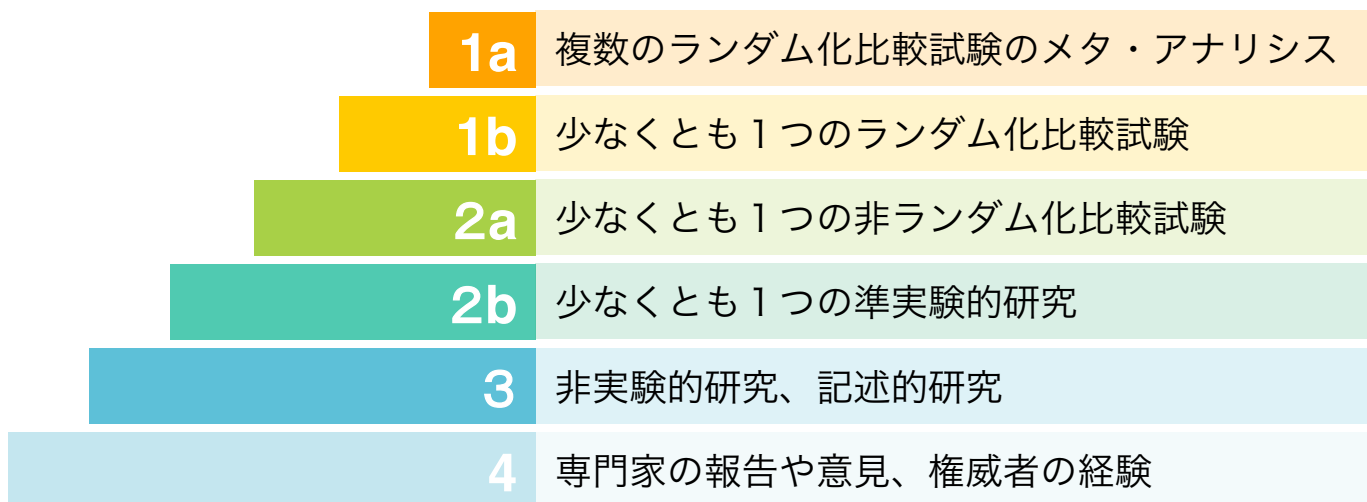


ランダム比較化試験に必要な要素

1. エンドポイント
施策の効果を計るための客観的な尺度
2. 比較対照
施策を実施する群（処置群）に対し、施策を実施しない群（対照群）
3. ランダム化
実施した施策以外の条件に偏りを生じさせないように、
処置群と対照群のランダムな振り分けを行う
4. 盲検化
計測する人に、処置群と対照群がどちらであるか分からないように
する（計測に主観が入らないようにするため）

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

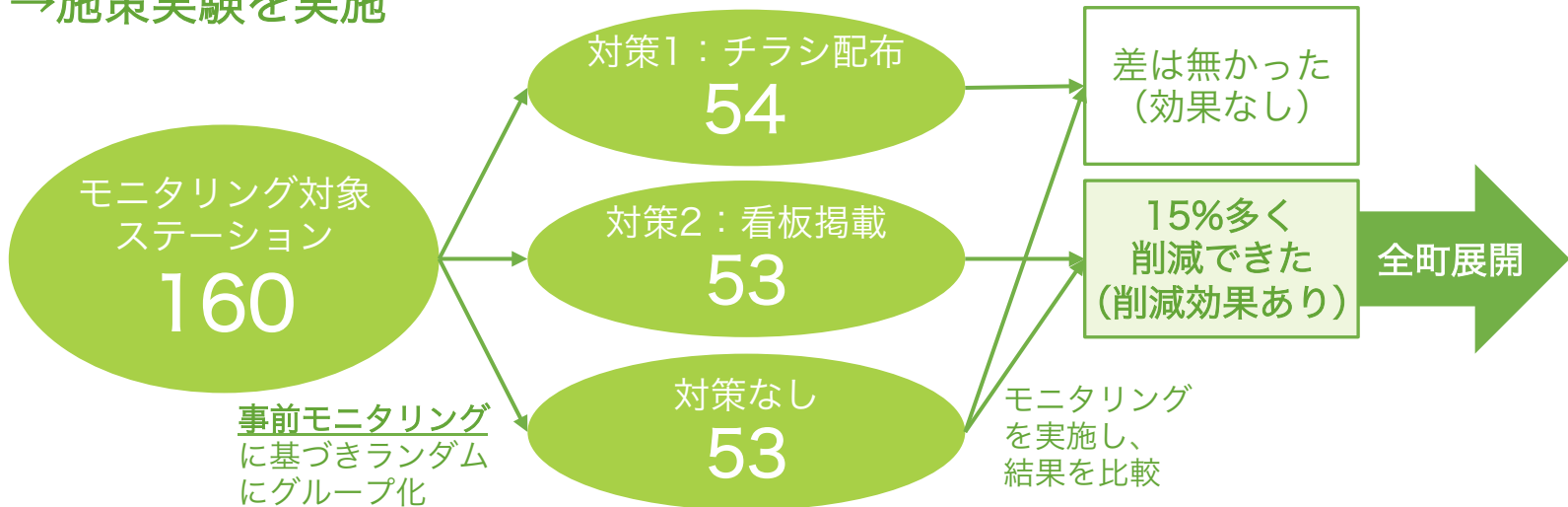


出典：正木朋也、津谷喜一郎（2006）「エビデンスに基づく医療（EMBの系譜と方向性：保健医療評価に果たすコクラン共同計画の役割と未来）」『日本評価研究 第6巻第1号』pp.3-20

事例：神奈川県葉山町の 放置ごみ削減施策の効果検証

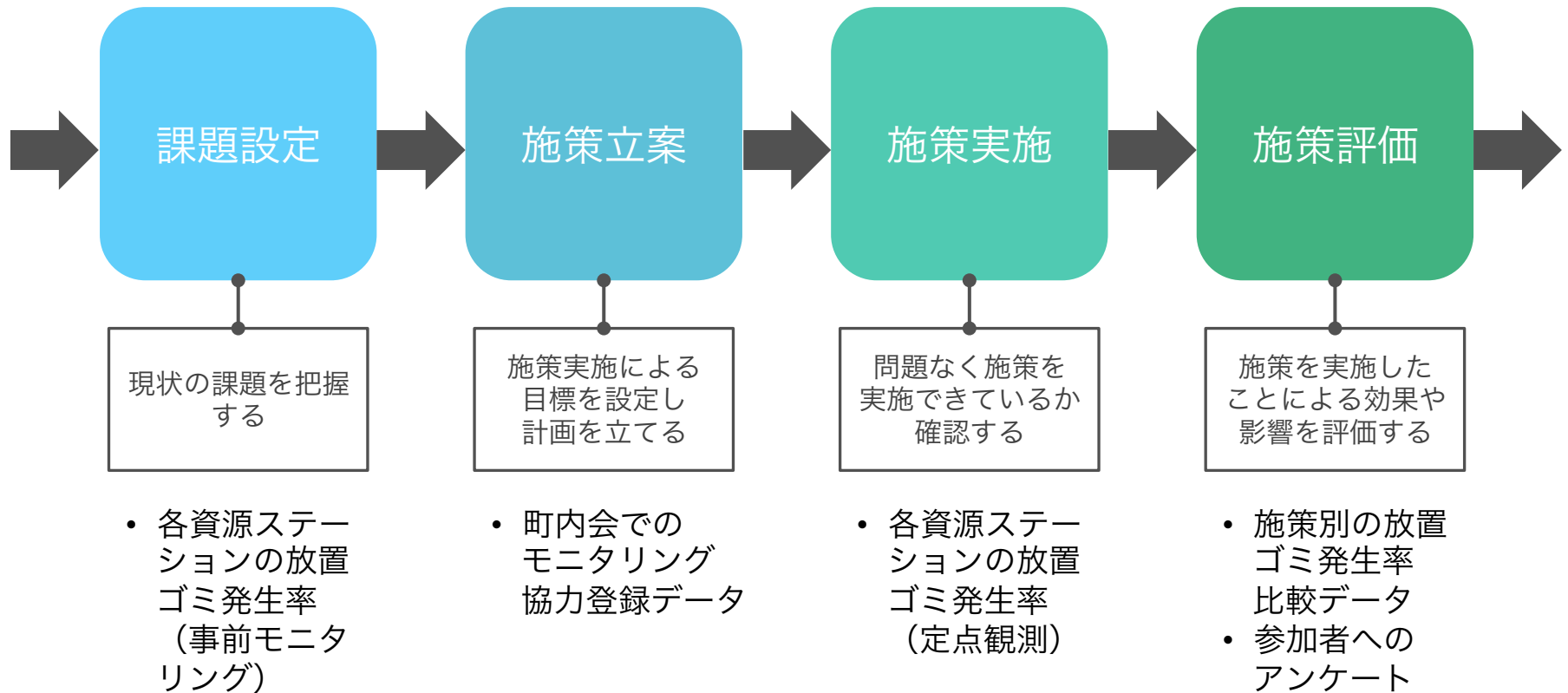
- 葉山町は2008年に日本で3番目に「ゼロ・ウェイスト」政策（ごみの徹底的な資源化・減量化）の導入を表明
- 資源ごみの収集拠点である資源ステーションの管理は、ゼロ・ウェイストの達成に向けた重要施策
- 資源ステーションでの放置ごみが減らないことに町・住民が頭を悩ませていた

→ 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/jp/events/bbl/16102501.html>

最終的に評価に必要なとなるデータを各プロセスにおいて取得

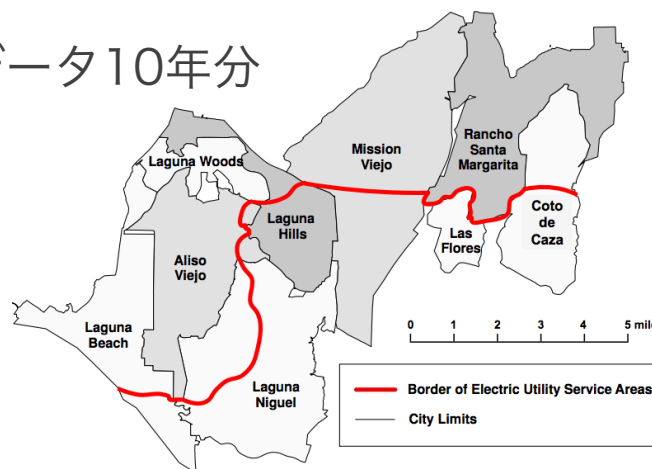


RCTが実施できない場合、質の高いエビデンスを得ることは不可能？

- 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 伊藤公一朗氏
- 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

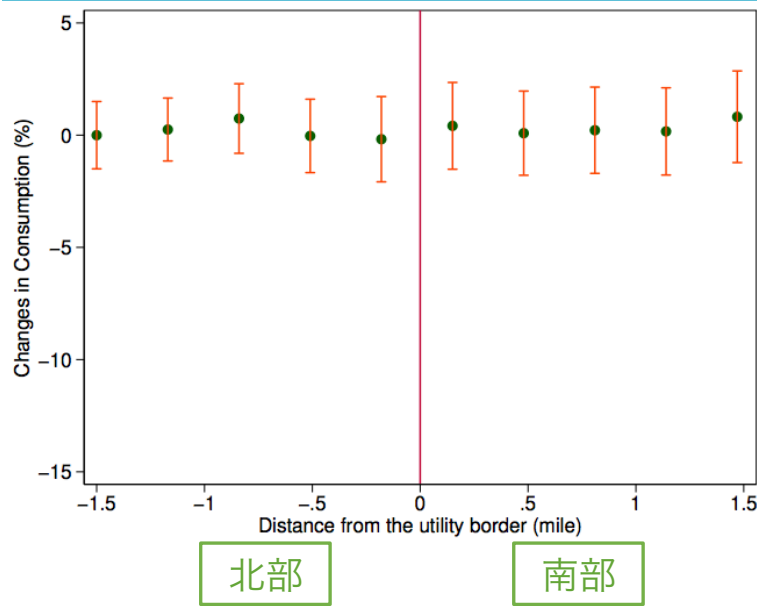


出典：政策の効果をどう測定するか？：海外における「エビデンスに基づく政策」の最新動向
<https://www.rieti.go.jp/jp/events/bbl/16102501.html>

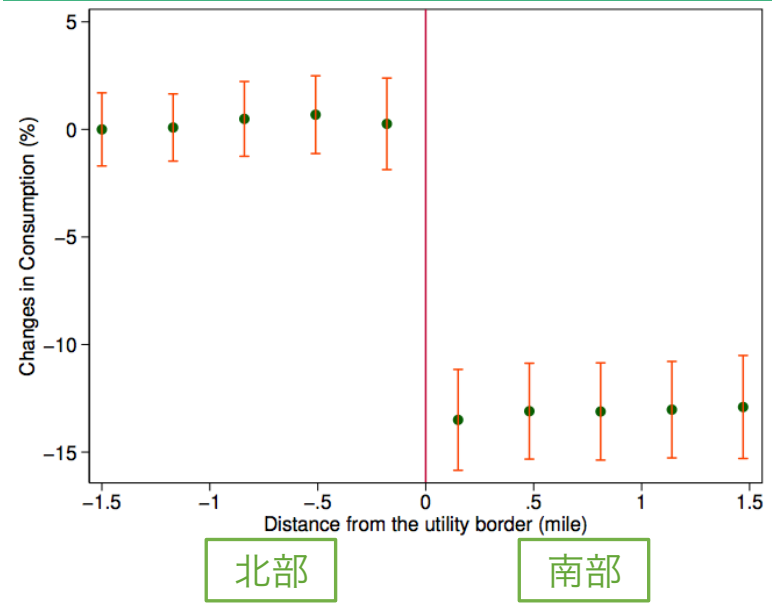
例) カリフォルニア州における電力消費量の分析

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

両地域で料金が同様だった時期



南部だけ料金が上がった時期



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うまく使い分ける
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科学的根拠として弱い場合がある

1. 行政データ分析を行う上で 覚えておきたい考え方（まとめ）

- 「データ」は「情報」を形式化したものであり、再度「情報」として解釈できる必要がある
- 情報は「データ化」することで分析が可能になり、新たな情報を取り出すことが可能になる
- 「定量データ」と「定性データ」はうまく使い分ける
 - 定量データ：数値で表すことができる、「俯瞰」のためのデータ
 - 定性データ：数値で表現できない、「深掘り」のためのデータ
- 相関があっても因果関係がない場合がある
- 因果関係の推定には「ランダム比較化試験 (RCT)」を用いるのが最も効果的
 - 施策を実施するグループ（処置群）と、施策を実施しないグループ（対照群）に分けて、指標を計測し比較
 - ランダムにグループ分けすることで、施策以外の要因による影響を排除できる
- 課題設定・施策立案・施策実施・施策評価の各プロセスにおいて、指標として使えるデータを取得することが重要

2.行政データ活用に立ちはだかる壁 とその乗り越え方

データ活用を進めるためには いくつかの壁がある

- いざ分析を進めようとした時に、律速要因になる
- どんな障壁が存在するのか、あらかじめ把握しておけばコストを見積もって計画を立てられる



①情報が特定の人の中 にしかない

- 特定の個人の経験や感覚に基づく情報しかない
- 例) ベテラン刑事の勘

- 再現性がない
- 他の人と共有できない

今夜は犯罪が
起こりそうだ



- まずはヒアリングやメモ等の定性データで良いので集める
- 判断の根拠としている情報を定量データ化できないか試みる

②データの所在が不明

- 目的のデータがどこの部課にあるか不明
- そもそもデータとして存在するのかわかっても不明

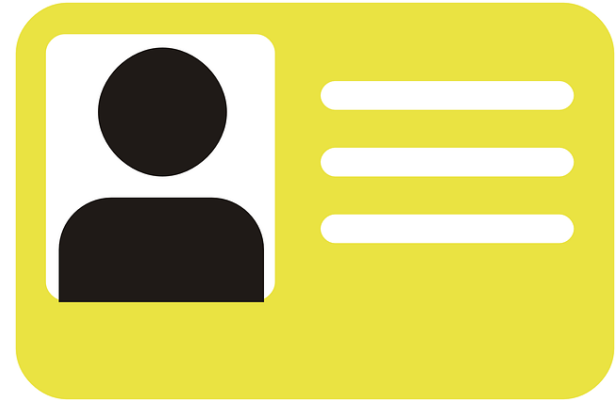
- 探す側も尋ねられる側も時間を浪費する
- すでにデータが他の部課にあるにも関わらず、同じような調査が行われたりして非効率



- 庁内でデータを共有する仕組みを作る
- まずは各課の所有データの目録を作って、どこに何のデータがあるか分かるようにする
- 調査を実施する時には、既に同様の調査が実施されていないか確認する

③データに個人情報が含まれている

- 個人情報保護条例等で、個人情報の含まれるデータの利用用途に制限が設けられている
- 部課間での受け渡しが禁止されている



- 分析目的によっては、統計処理で個人情報が含まれない状態にして利用可能
- 個人情報保護条例において目的内利用に該当するか等を確認の上、方法を検討する（後述）
- 個人情報の活用によって住民の生活を向上させるという面も念頭に置いて、条例を改正



2017年 個人情報保護法の改正

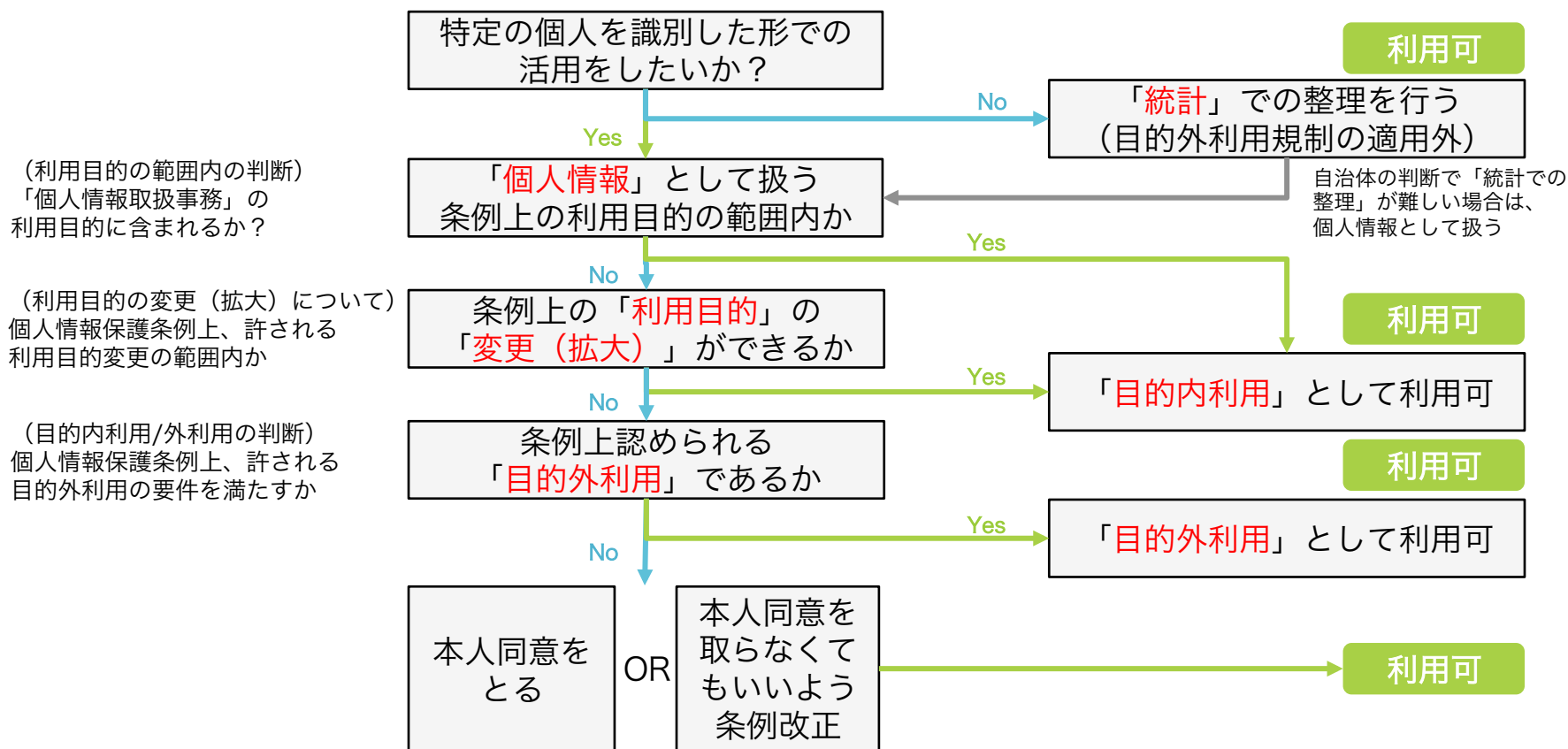
個人情報を社会で活用可能にするという観点で方針が定められた

第一章 総則

(目的)

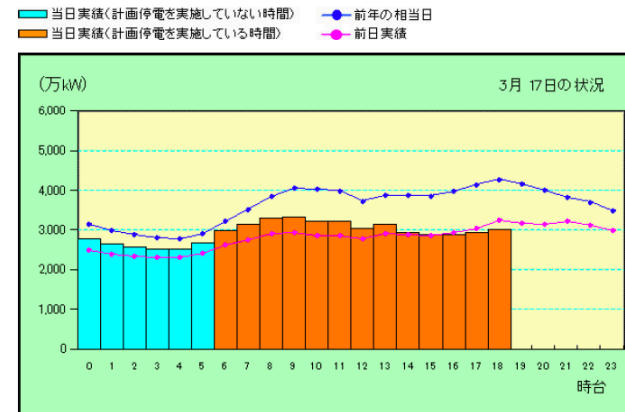
第一条 この法律は、高度情報通信社会の進展に伴い個人情報の利用が著しく拡大していることに鑑み、個人情報の適正な取扱いに関し、基本理念及び政府による基本方針の作成その他の個人情報の保護に関する施策の基本となる事項を定め、国及び地方公共団体の責務等を明らかにするとともに、個人情報を取り扱う事業者の遵守すべき義務等を定めることにより、個人情報の適正かつ効果的な活用が新たな産業の創出並びに活力ある経済社会及び豊かな国民生活の実現に資するものであることその他の個人情報の有用性に配慮しつつ、個人の権利利益を保護することを目的とする。

個人情報利用チェックフロー



④データがコンピュータで扱える形式になっていない

- 紙媒体でしか保管されていない
- 画像でしか保管されていない

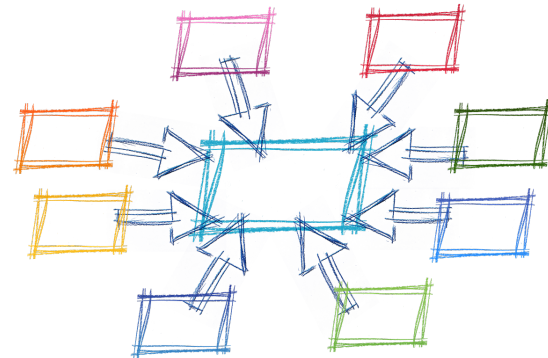


- 他の部課や外注先に元の生データがないか探す
- 重要な情報で生データが無い場合は、外注等も活用しながらテキストデータ化し、庁内で共有できる形で保管する
- 調査委託等では生データも納品するような契約になっているか見直す

⑤情報をデータ化する過程で重要な情報が抜け落ちてしまっている

- 表などの形式に当てはめようとして情報を削り過ぎてしまう
- 数値情報を扱う時に、集計後のデータしか保管していない例)
アンケート調査データ

- データから重要な情報を読み取れない
- 分析に使えない



- 定性データもちゃんと残す
- アンケートや統計調査を行なったら、集計データだけでなく
個票データも保管する

⑥ データから情報が正しく 取り出せない

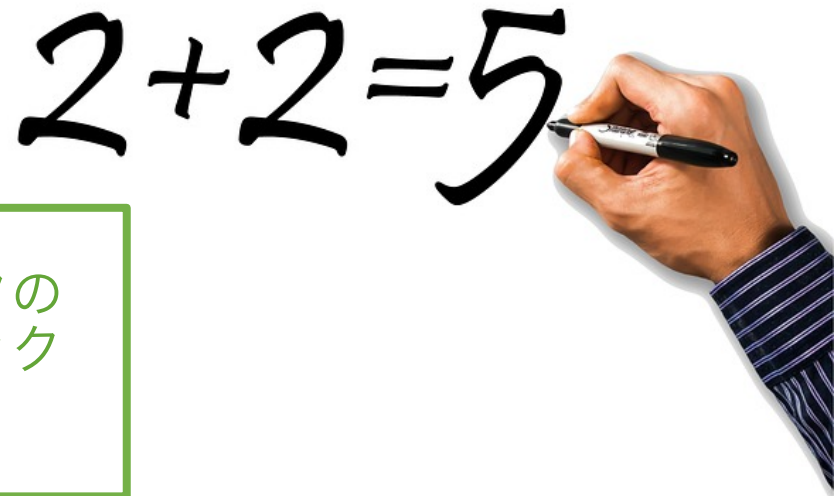
- 情報が正しく取り出せない状況のまま分析やサービス開発を進めてしまう
 - 背景知識を持たずにデータを使ってしまう
 - 誤った手法で分析してしまう
 - 結果、情報の解釈の仕方を間違えて、使えない分析結果やサービスになる

- 背景知識 + 分析知識を持って取り組む
- 部課を越えて連携体制を構築する
- 企業・大学・研究機関等とチームを組む



⑦データに間違いが含まれたまま気付かない

- 間違いが含まれた状態でデータが使用されてしまう
 - 元のデータから間違っているケース
 - 加工する過程で何らかのミスが発生し、そのまま残ってしまうケース



人は必ず間違っているので、データの設計やシステムによるチェックで回避する方法を検討

データ活用を進めるためには いくつもの壁がある

- ① 情報が特定の人の中だけにしかない
- ② データの所在が不明
- ③ データに個人情報が含まれている
- ④ データがコンピュータで扱える形式になっていない
- ⑤ 情報をデータ化する過程で重要な情報が抜け落ちてしまっている
- ⑥ データから情報が正しく取り出せない
- ⑦ データに間違いが含まれたまま気付かない

これらの他にも壁はあるかもしれません。
対処法を把握して乗り越えましょう！

3. データ分析の流れと分析手法の選び方

生データをそのまま見ても 良く分からない

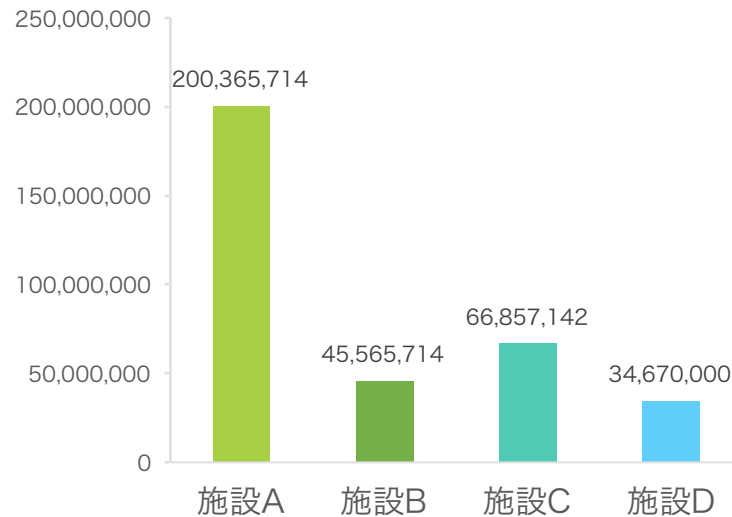
例：ある市の指定管理施設の年間の利用状況

施設名	開館日数 (日)	利用者数 (人)	利用料（観覧 料等）（円）	指定管理料 (円)
施設A	331	530,776	157,018,640	200,365,714
施設B	285	29,737	5,358,050	45,565,714
施設C	303	112,066	34,534,080	66,857,142
施設D	297	27,125	12,893,260	34,670,000

可視化すると 状況が分かりやすくなる

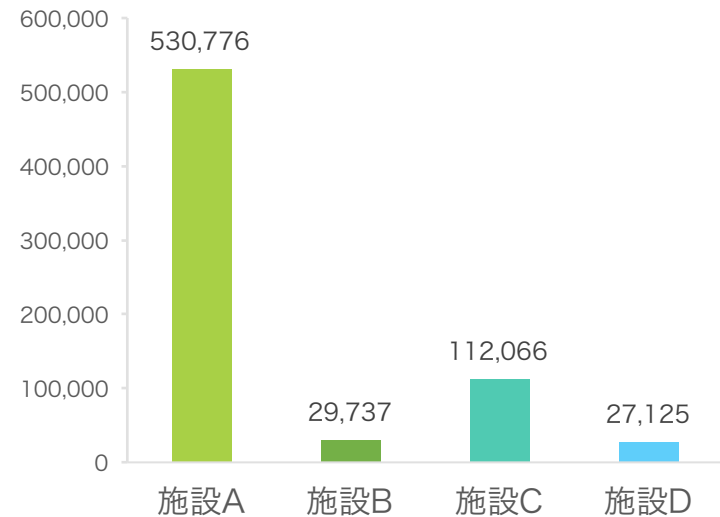
例：市の指定管理施設の年間の利用状況

指定管理料（単位：円）



施設Aの運営に多くのコストが割かれている

年間の利用人数（単位：人）



コストがかけている施設は利用者が多い

分析→可視化すると さらに詳しい状況が見えてくる

例：市の指定管理施設の年間の利用状況

各施設のコストパフォーマンス
を知りたい



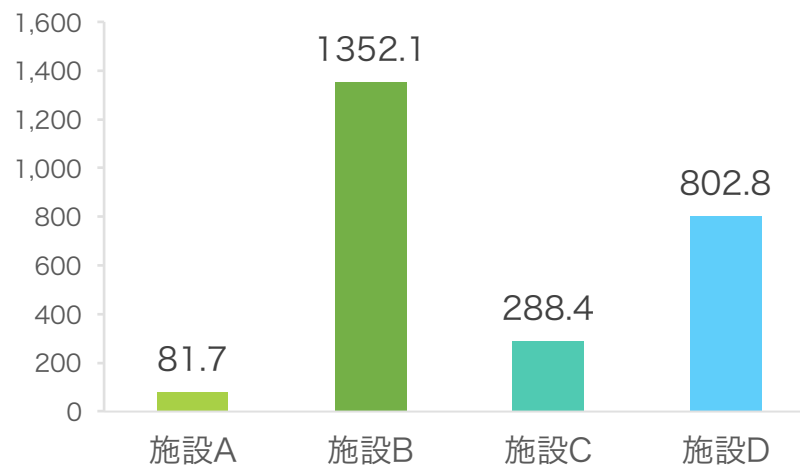
利用者1人あたりにサービスする
ためにかかる金額を計算

(指定管理費 - 利用料)

÷

利用人数

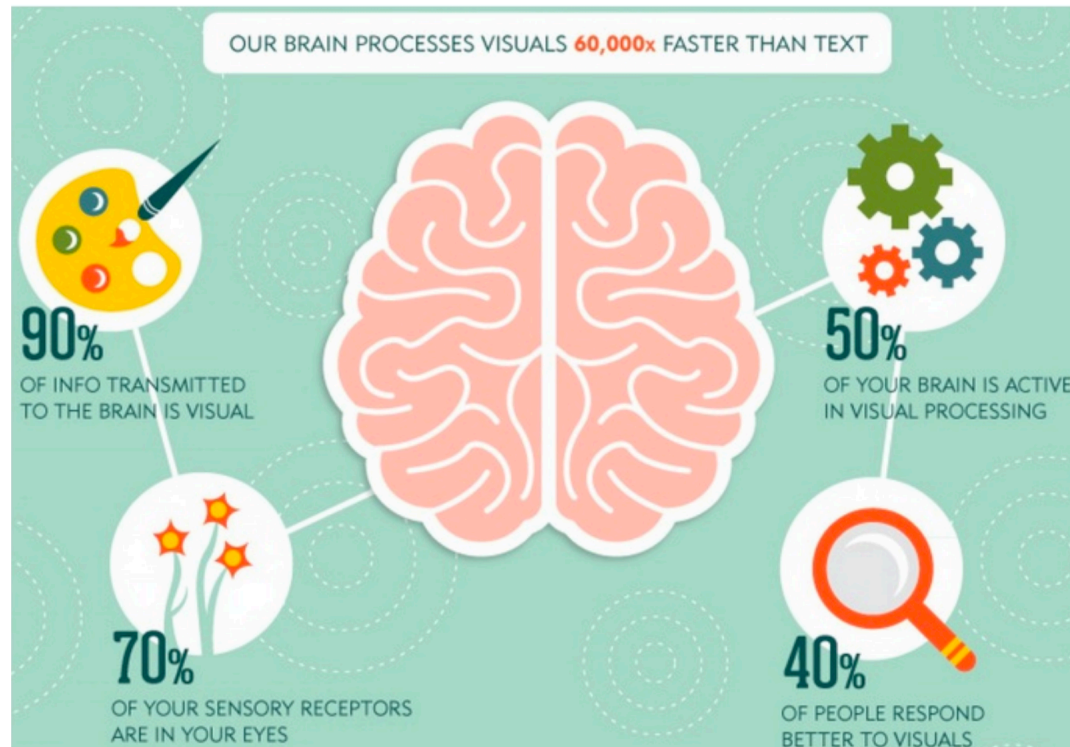
1人にサービスするためにかかる金額
(単位：円)



施設Aの運営には最も多くの予算が割
かれているが、コストパフォーマンス
は一番高い

人間の脳はビジュアルを テキストの6万倍速く認知

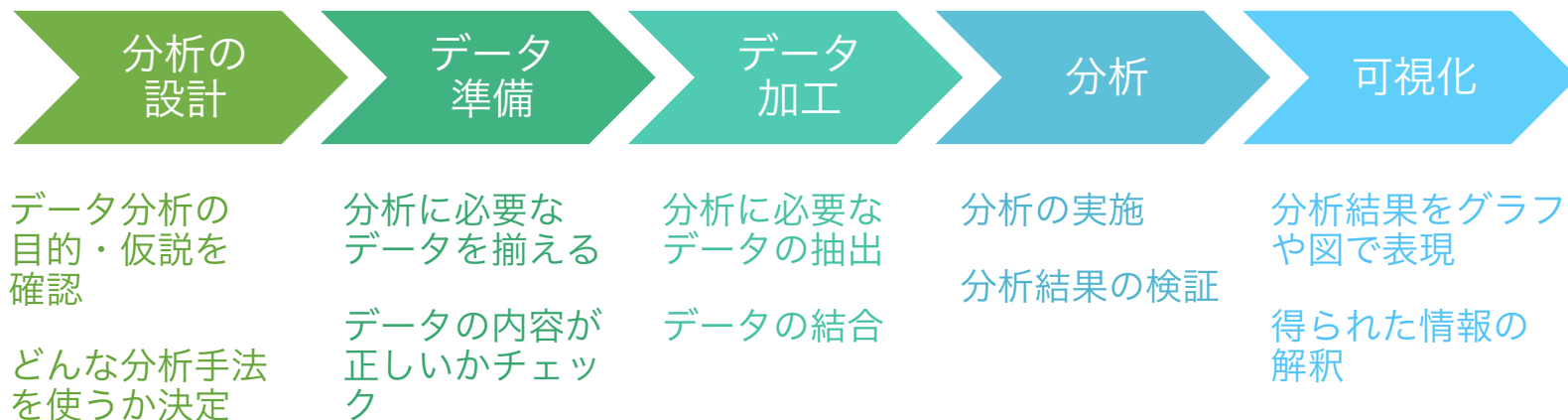
数字だけ見せるよりも、ビジュアライズすると伝わりやすくなる



<https://www.slideshare.net/elsekramer/show-dont-tell-the-rise-of-visual-on-social-media/35-brand-identity>

データ分析のプロセス

- データ分析とは、データの中から有用な「情報」を見つけ出し、意思決定に使えるようにするためのプロセス
- データ分析を行う「目的」「仮説」をまず定めておくことが重要
 - 何を明らかにしたいのか？
 - 誰が、どのような意思決定に使うのか？



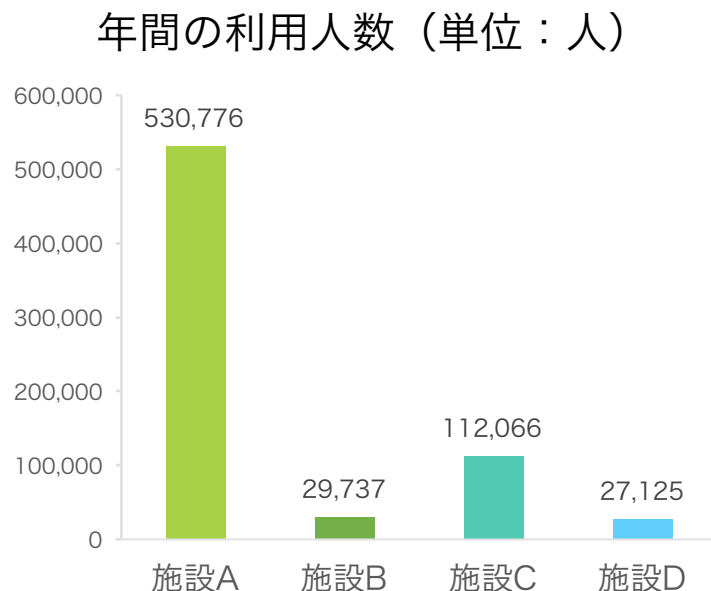
「グラフ」は世界中でよく使われるデータの可視化手法



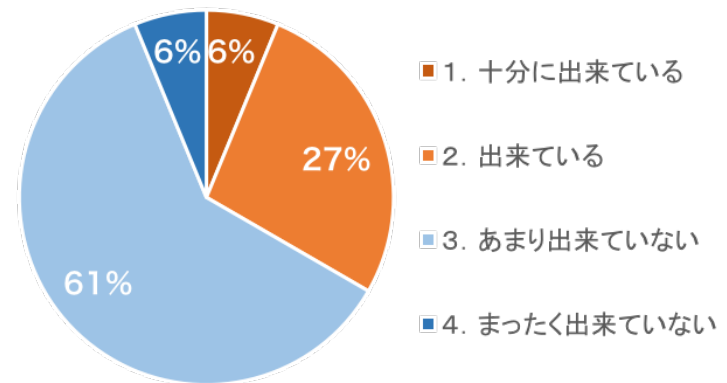
By Mmbcn [CC BY-SA 3.0 (<https://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons

グラフでできること： 1. 大小を比較する

グラフでデータ項目同士の数値や割合を比較することで、違いが見つけやすくなる



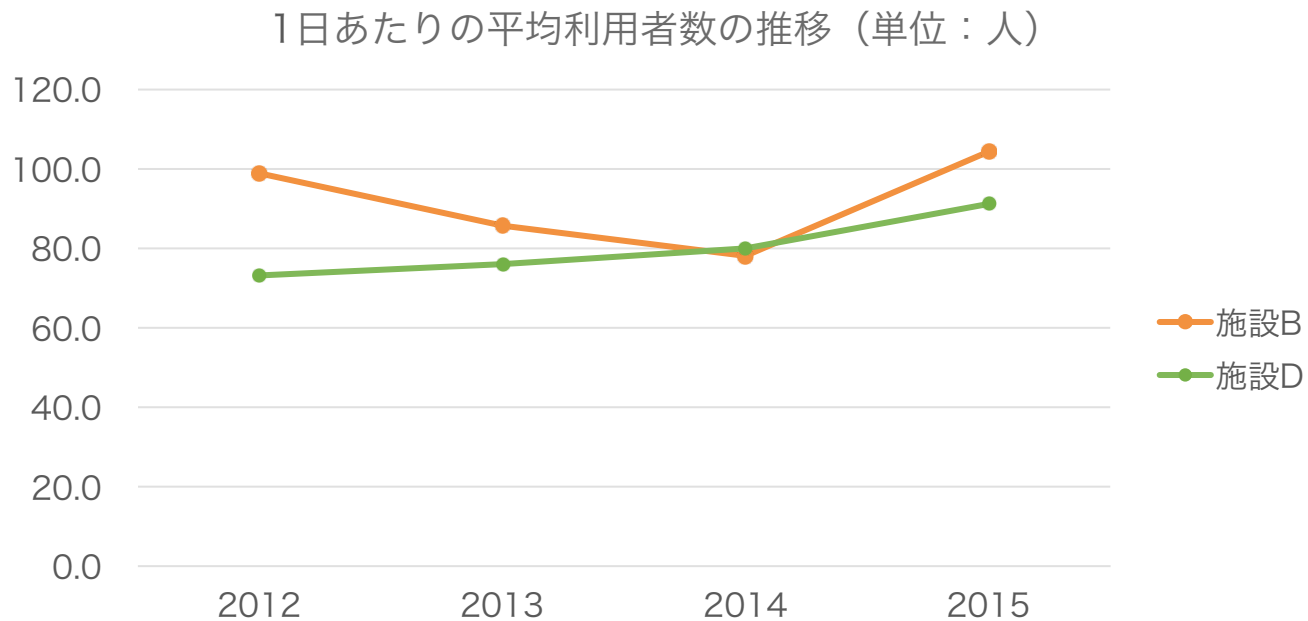
アンケート結果：普段の業務でデータ活用はできていますか？



グラフでできること：

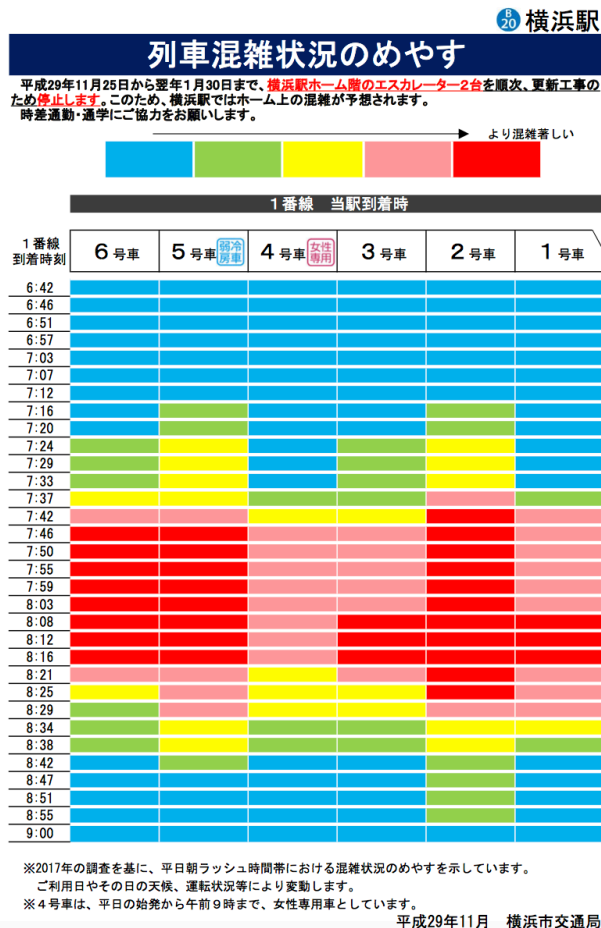
2. 傾向を把握する

グラフで時系列のデータを比較することで、時間の推移によるデータの変化の傾向を把握することができる



グラフでできること：

3. 偏りを見つける

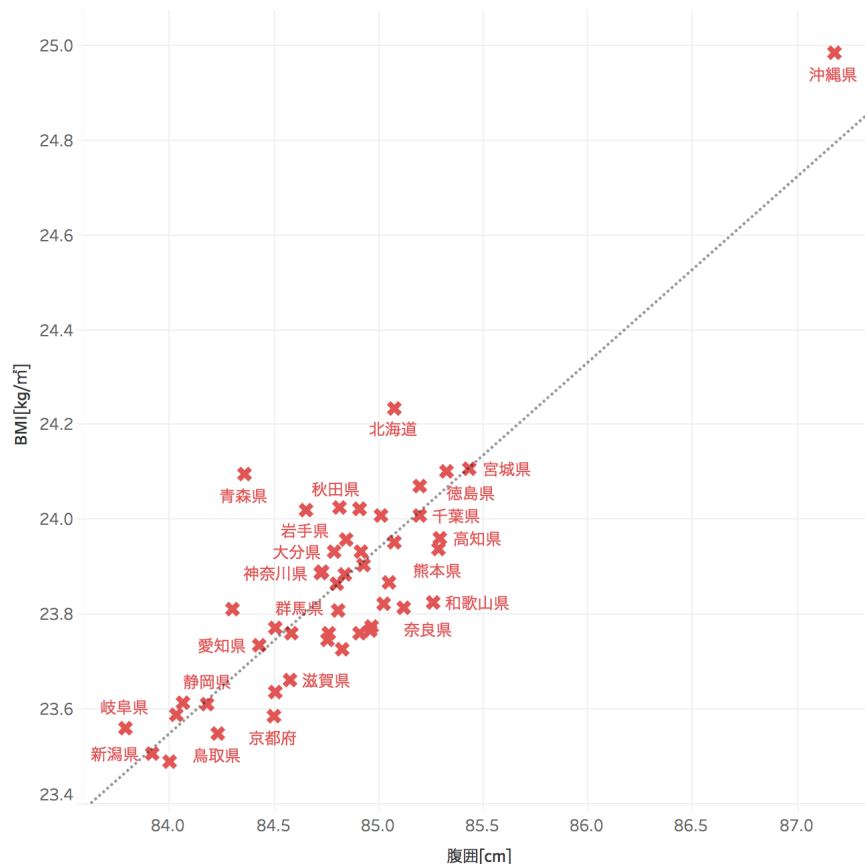


数値の大きさに応じて色分けしたりマークの大きさを変えることで、データの偏りを見つけやすくなる

例) 横浜市営地下鉄の混雑状況のヒートマップ

グラフでできること： 4. 関係性を見つける

特定健診結果のBMIと腹囲の相関（都道府県別平均値・男性）



データ項目間に相関関係があるかどうかを見つける

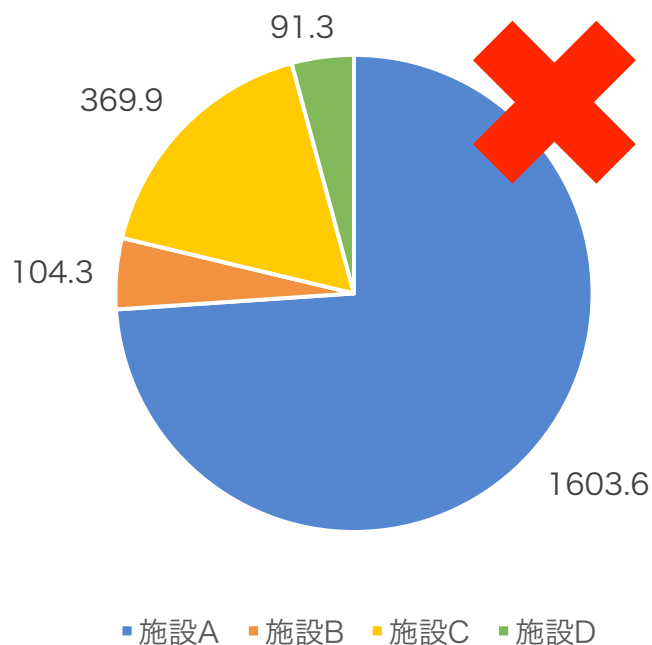
例) 特定検診結果のBMIと腹囲の散布図
(都道府県別平均値・男性)

<https://public.tableau.com/views/BMI/BMI>

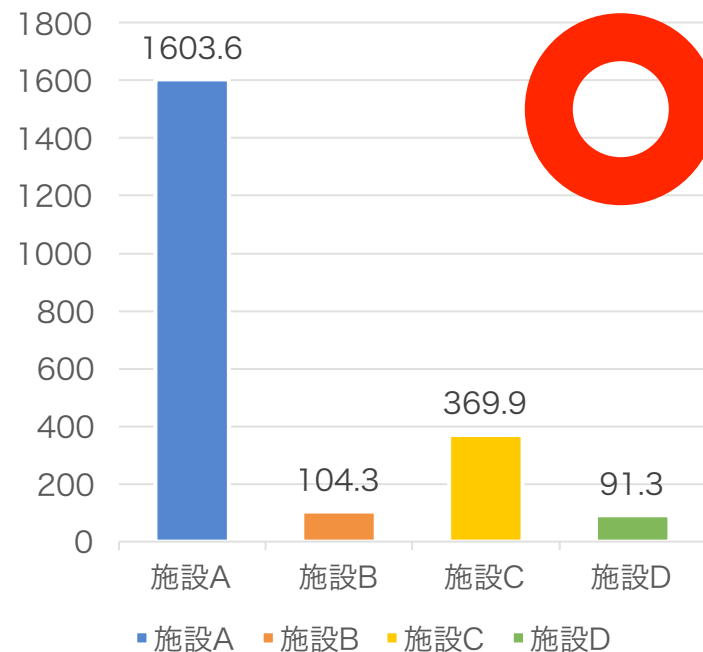
目的によって適切/不適切なグラフがある

Excelなどのツールを使うと、適当に選ぶだけでそれっぽいグラフを作成できてしまうので要注意

1日あたりの平均利用者数 (2015年)



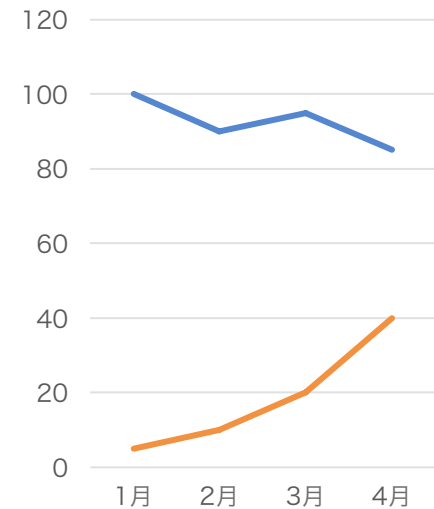
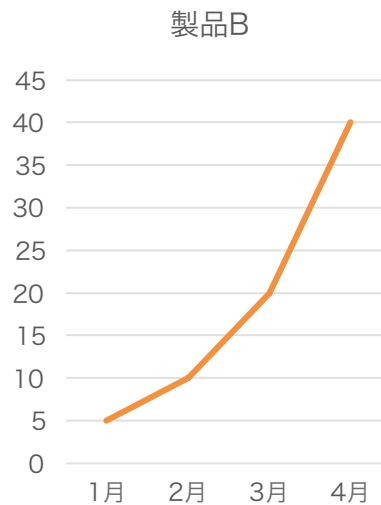
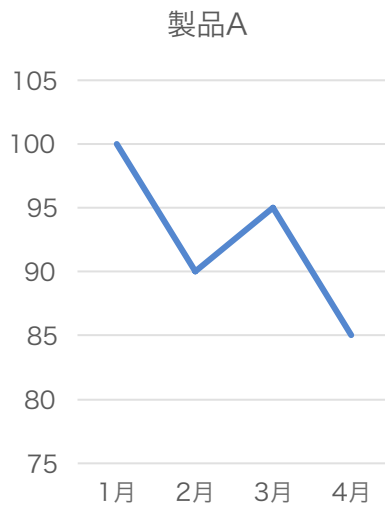
1日あたりの平均利用者数 (2015年)



世の中にはかなり恣意的に表現されているグラフも存在する

都合の良いデータだけ表示したり、軸をあえて統一しなかったり

例) 製品の売り上げの比較



基本グラフの 目的別選択チャート

データの大きさを
比較したい

量の比較



棒グラフ

割合の比較



円グラフ

時間の経過による推移を知りたい



折れ線グラフ

データの偏りを知りたい



ヒートマップ

データ項目間の関係を知りたい



散布図

※必ずしもこれに当てはまるとは限らないので目安として

GISを用いた分析

GISとは

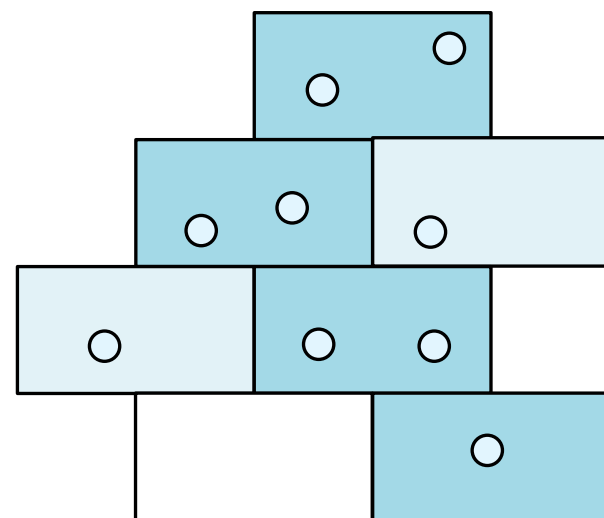
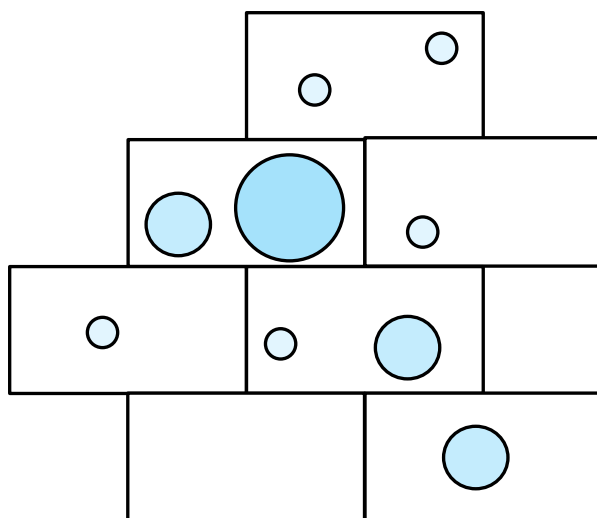
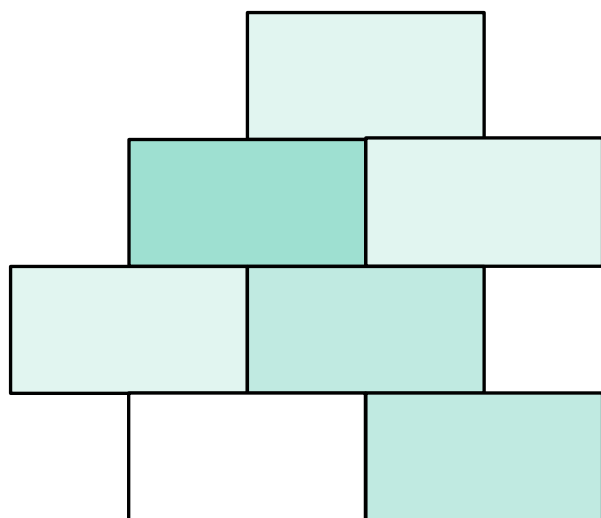
- 地理情報システム（GIS：Geographic Information System）は、地理的位置を手がかりに、位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術。
（出典：国土地理院）
- 地図上にデータを重ねて分析することで、対象の地理空間的な特性を明らかにすることができる
- 地図上に表示することで、データを直感的に把握することができる

GISを用いた分析の例

種別や割合
ごとに色分け

回数や頻度
による表示

ポイント数
による表示



分析の目的や、誰に何を見せたいのかを考慮して手法を選択する

GISを用いた分析の具体例

GISを使った基本的な分析方法

■ 行政区域のポリゴン（面データ）を色分けして分析

- 高齢者率や、高齢者人口のデータを用いてポリゴンの色分けをして、高齢化の進んでいる地域を確認する
- 近接する地域でどのような傾向があるか、河川や山間部などの地理的な位置との関係で傾向があるか等を確認する

■ レイヤーを重ね合わせた分析

- 福祉施設と、地区ごとの高齢者の分布ポリゴンを重ね合わせて、適切な配置ができているか確認する
- 避難所のデータと、人口データを重ね合わせて、避難所のキャパが足りているか確認する

4. 公共データの活用が もたらすインパクト

データの組み合わせは新たな価値を生む： サンフランシスコ市 × Yelp



- 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



データの組み合わせは新たな価値を生む： サンフランシスコ市 × Yelp

yelp Find Restaurants Near San Francisco, CA, United States Log In Sign Up

Home Services Restaurants Auto Services More Write a Review

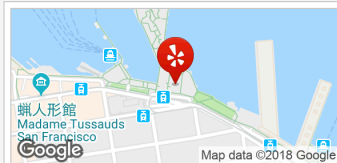
Fog Harbor Fish House Claimed

4550 reviews Details

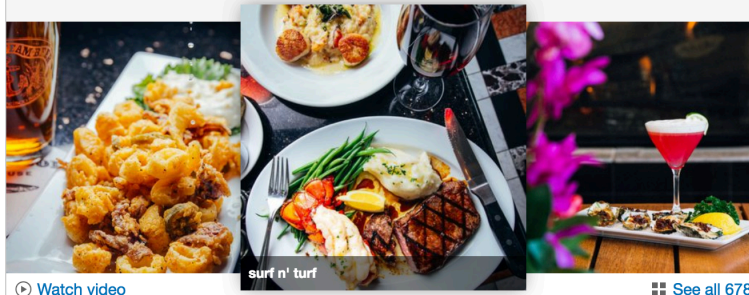
Write a Review

Add Photo Share Save

Seafood, Bars Edit



Pier 39
Ste A-202
San Francisco, CA 94133
Fisherman's Wharf, North Beach/Telegraph Hill
Get Directions
(415) 421-2442
fogharbor.com
Make a Reservation
Send to your Phone



Watch video

See all 6788

Skip the line, Join our waitlist via your phone!

Join Now



"I've taken numerous visitors to the wharf for **clam chowder**, sourdough bread, and a ton of salt water taffy." in 1394 reviews
\$7 Award Winning Clam Chowder



"We also loved the **blue cheese garlic bread** (I know, sounds strange but is fantastic) dipped in the red curry steamed mussels broth." in 88 reviews



"The **crab cake** was delicious and filling and the garlic roast crab was as good as you can find on the wharf." in 298 reviews
\$19 Dungeness Crab Cakes

Show more review highlights

Ask the Community

Make a Reservation

Thursday, September 13, 2018

7:00 pm

2 people

Find a Table

Today 11:00 am - 10:00 pm
Closed now

Full menu

Price range \$11-30

Health Score 87 out of 100

衛生スコア：
87点

データの組み合わせは新たな価値を生む： サンフランシスコ市 × Yelp

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？

炎上、ヘタすると訴訟

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

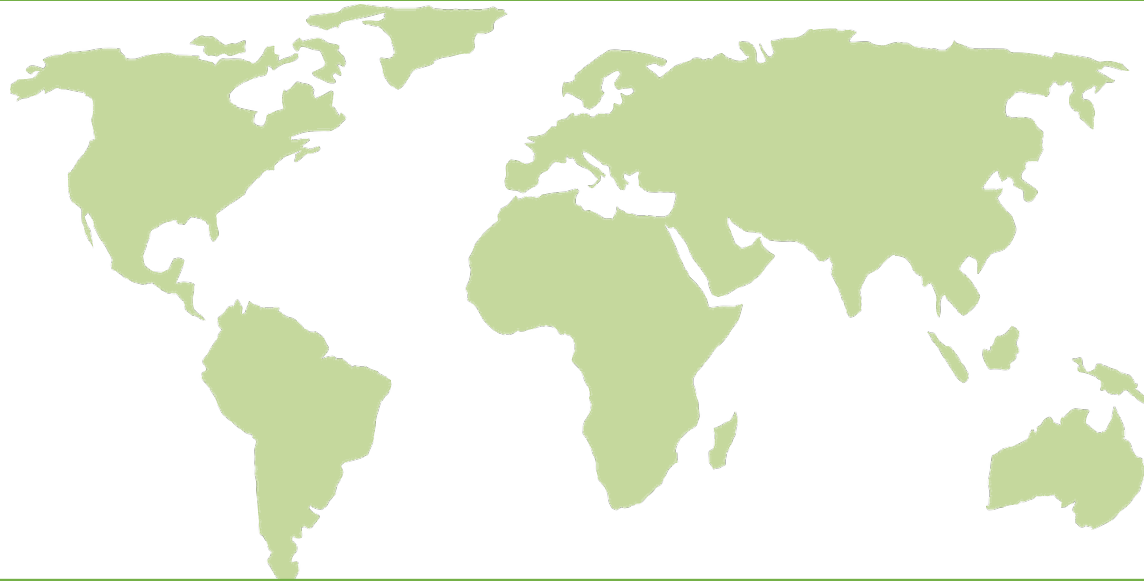
国、地方公共団体及び事業者が保有する官民データのうち、国民誰もがインターネット等を通じて容易に利用（加工、編集、再配布等）できるよう、次のいずれの項目にも該当する形で公開されたデータをオープンデータと定義する

【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



2016年 官民データ活用推進基本法

国と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

オープンデータ基本指針 概要

本基本指針の位置づけ

平成28年12月14日に公布・施行された「官民データ活用推進基本法」において、国、地方公共団体、事業者が保有する官民データの容易な利用等について規定された。本文書は、これまでの取組を踏まえ、オープンデータ・バイ・デザイン^(注)の考えに基づき、国、地方公共団体、事業者が公共データの公開及び活用に取り組む上での基本方針をまとめたものである。

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

2. オープンデータの定義

- 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

3. オープンデータに関する基本的ルール

- (1) 公開するデータの範囲・・・各府省庁が保有するデータは、原則オープンデータとして公開。公開することが適当でない公共データは、**公開できない理由を原則公開**するとともに、限定的な関係者間での共有を図る「**限定公開**」といった手法も積極的に活用。
- (2) 公開データの二次利用に関するルール・・・原則、**政府標準利用規約**を適用。
- (3) 公開環境・・・特にニーズが高いと想定されるデータは、一括ダウンロードを可能とする仕組みの導入や、APIを通じた提供を推進。
- (4) 公開データの形式等・・・**機械判読に適した構造及びデータ形式で掲載**することを原則。法人情報を含むデータは、**法人番号を併記**。
- (5) 公開済みデータの更新・・・可能な限り**迅速に公開**するとともに**適時適切な更新**。

4. オープンデータの公開・活用を促す仕組み

- (1) オープンデータ・バイ・デザインの推進・・・行政手続き及び情報システムの**企画・設計段階から必要な措置**
- (2) 利用者ニーズの反映・・・各府省庁の保有データとその公開状況を整理した**リストを公開**→利用者ニーズを把握の上、**ニーズに即した形での公開**

5. 推進体制

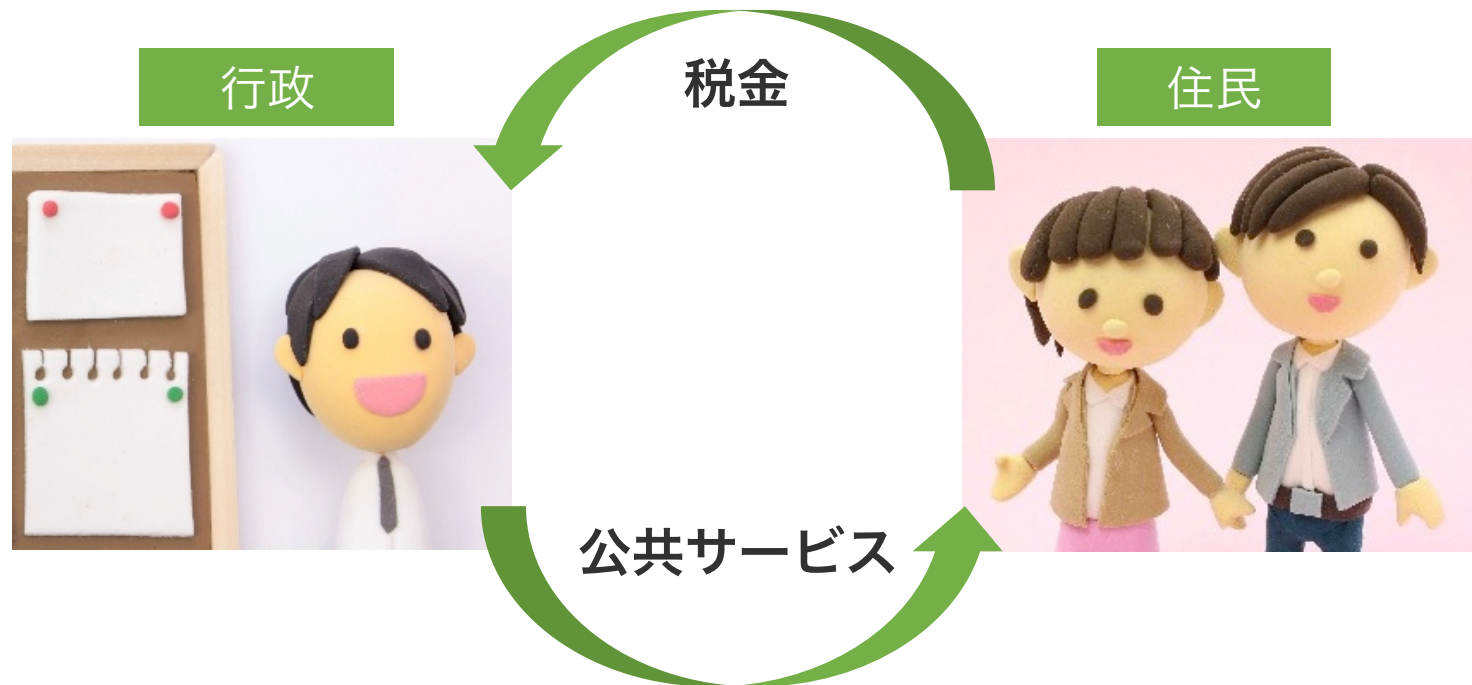
- (1) 相談窓口の設置・・・総合的な相談窓口（内閣官房IT総合戦略室）・相談窓口（各府省庁）の設置
- (2) 推進体制・・・内閣官房IT総合戦略室は、政府全体のオープンデータに関する企画立案・総合調整、各施策のレビュー、フォローアップを実施等

6. 地方公共団体、独法、事業者における取組

地方公共団体・・・官民データ法の趣旨及び本基本指針を踏まえて推進。
独立行政法人・・・国費によって運営されていること又は実施している事業や研究があることに鑑み、基本指針に準拠して取組を推進することが望ましい。
公益事業分野の事業者・・・その公益性に鑑み、本基本指針及び利用者ニーズを踏まえて推進することが望ましい。

(注) 公共データについて、オープンデータを前提として情報システムや業務プロセス全体の企画、整備及び運用を行うこと。

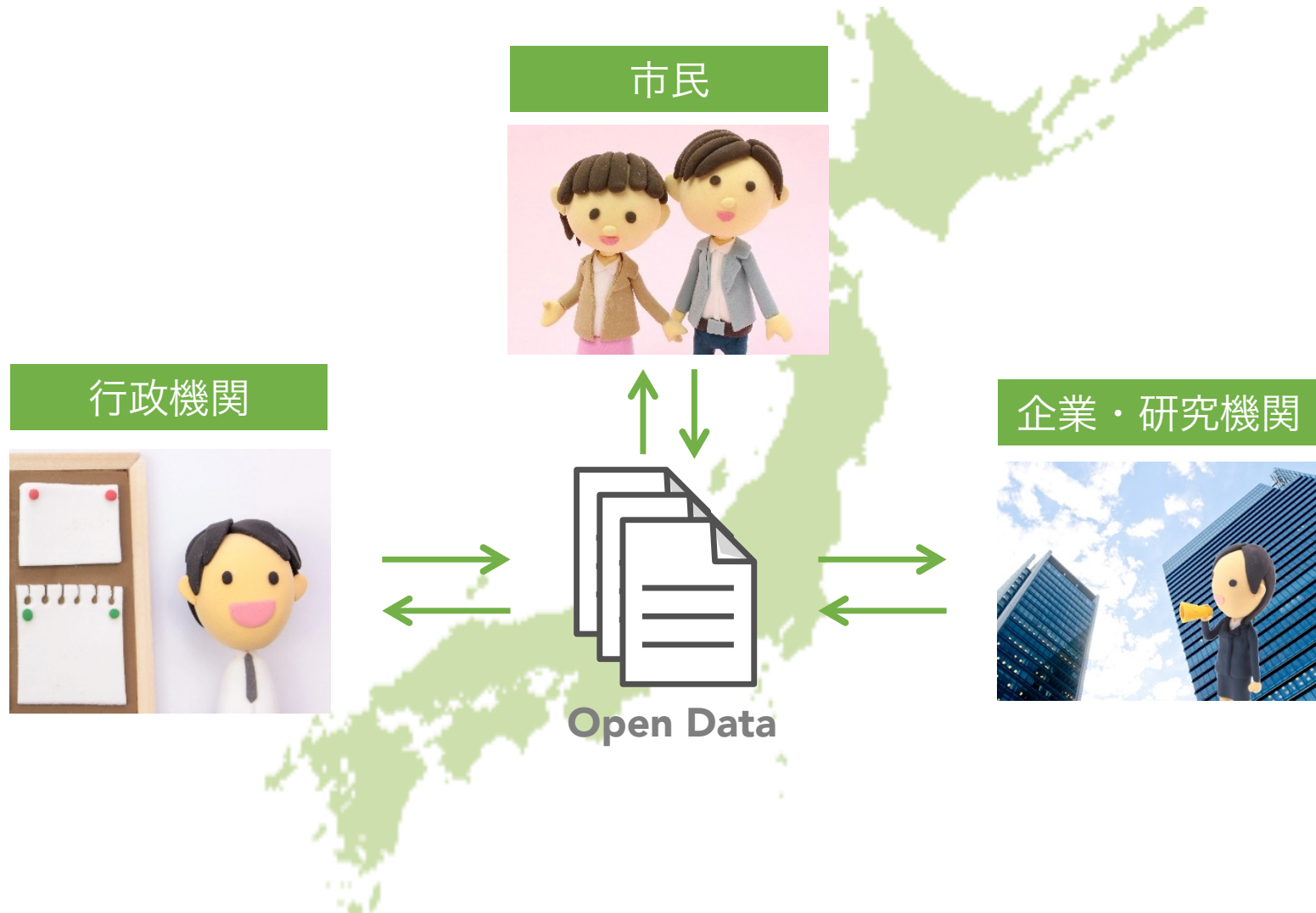
これまででは行政が公共サービス運営の主体だった



行政職員だけで検討・対応するには限界が来ている



オープンデータによって実現する 組織を超えて得意分野を分担できる社会



事例：「しずみちInfo」 (静岡市 × トヨタ自動車)

静岡市「しずみちInfo」

- 道路交通情報をオープンデータ化
- 通行規制や災害発生などのデータをリアルタイムで配信

トヨタIT開発センター

- 共同実験で「しずみちInfo」のデータを取得しカーナビで情報配信

更新時刻	規制区分	道路名	規制開始日	規制原因	規制種別
2018/09/11 17:06	工事	井川湖御幸線	18/09/12 08:30	工事(復...	時間帯全面通行止
2018/09/10 09:36	工事	橋北脇新田線	18/09/12 08:30	工事(道...	片側交互通行
2018/09/08 06:30	工事	国道136号	18/09/09 22:00	工事(道...	時間帯全面通行止
2018/09/05 18:30	事前通行規制	473号	18/09/04 21:30	雨量	その他規制
2018/09/04 18:30	災害差	国道472号	18/09/04 11:00	災害差	全面通行止

- 行政が一番早く情報を入手していたり、行政しか持っていないデータもある
- 全てのサービスを行政が提供するの難しいので、民間が得意な所は民間に任せる

事例：「ちばレポ」 千葉市における市民協働

■ 市内で起きている課題を、市民がサイトに投稿することでレポートできる

- 「公園の遊具が壊れている」
- 「道路がへこんでいて危険」

■ 課題は地図上で可視化され、行政側からは対応ステータスが情報提供される

- 「レポートを受け付けました」
- 「舗装の補修が完了しました」

ちばレポ ちば市民協働レポート

ちばレポとは? ログイン レポーター・サポーター登録 文字サイズ 小 中 大 千葉市 CHIBA CITY

トップ レポートを見る/する サポーター活動 アイデア提案

「こまった」レポート テーマレポート かいけつレポート トップページ > レポートを見る/する

直近に更新のあった200件の「こまった」レポートを表示しています。[地図表示 | 一覧表示]

レポート条件検索 このページの見方 レポートガイド 公開ガイドライン 地図設定 レポート開始

例: 千葉市中央区千葉港1-1 検索

レポートNO: No-20180729-223196

レポート日時: 2018/07/29 18:30

対応状況: 対応済

対応済

分野: 道路

ニックネーム: ロッキー

ロッキー

タイトル: 道路がへこんでいて危険

道路がへこんでいて危険

閉じる レポート詳細へ

(C)Yahoo Japan.(C)ZENRIN

- 市民は自分の生活している周辺の状況変化をいち早く察知
- 行政は効率良く地域の課題を把握し、対応できる
- 課題がデータとして共有されることで、市民の自治意識が高まる

(付録) 庁内で実践を進めたい方に おすすめのプログラム

地域情報化アドバイザー派遣制度

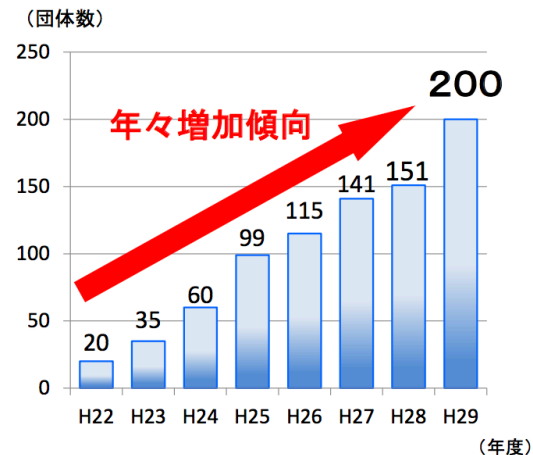
http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/manager.html

- 地域が抱える様々な課題を解決するため、ICTを活用した取組を検討する地方公共団体等からの求めに応じ、ICTの知見等を有する「地域情報化アドバイザー」を派遣し、ICT利活用に関する助言等を行う。

派遣の仕組み



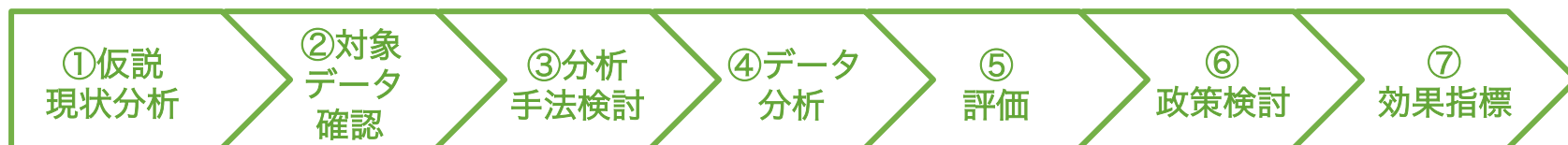
派遣団体数



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行い、11の自治体にて実証実験を行った（2018年度も実証継続中）
- 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
- 7stepを通常半日×4日間の研修で習得する



データアカデミーに関するお問い合わせ先：data-academy@code4japan.org



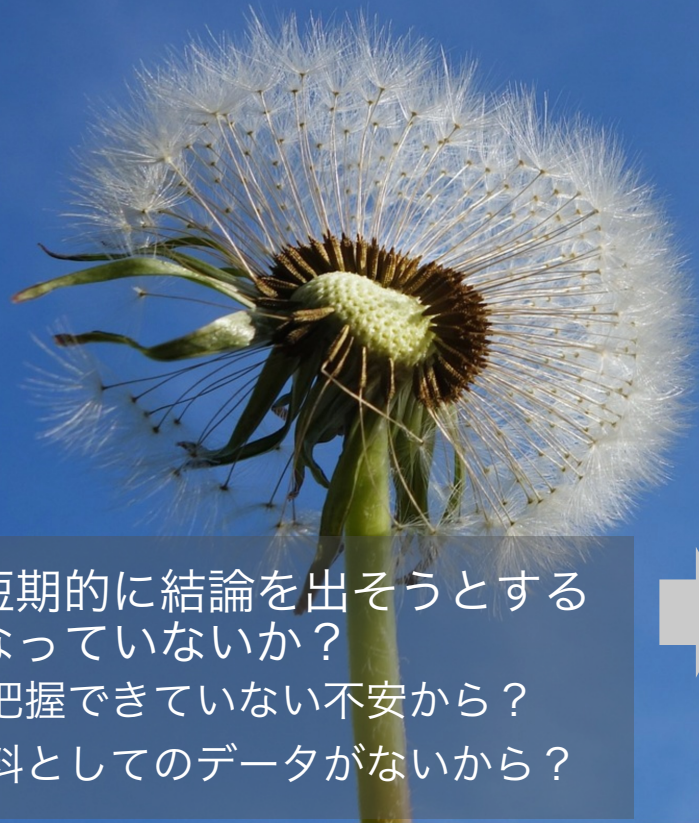
データアカデミーの特徴

比較項目	一般的なデータ分析研修	Code for Japanのデータアカデミー
研修対象	庁内データを使った統計・GIS分析研修	庁内データ利活用のためのプロセス研修
研修課題	他市の事例や、一般的事例をトレースする	現課から提出された 実際の課題 を利用する
自治体の規模	大きな自治体で行う	政令市から町村まで対応可能
方法	先生・講師型、座学型の集合研修	複数の課が参加する アクティブラーニング 研修

個別のデータ分析技術を覚えるのではなく、データ分析を課題解決プロセスとして利用できるスキルを身につける

データアカデミーに関するお問い合わせ先：data-academy@code4japan.org

長い目で見通せる社会へ



現状が
把握できる

将来の
見通しが
立てられる

長期的に
物事を
考えられる

現代は短期的に結論を出そうとする社会になっていないか？

- 現状が把握できていない不安から？
- 判断材料としてのデータがないから？

もし判断に必要なデータと分析手段が全部入手できるようになったら？

Link Data Now!!



LinkData.org

本資料に関するご質問・ご意見や、
データ活用に関するご相談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