

一般講座

## 「IoT時代のデータ活用」WEBセミナー

# DX推進におけるデータ活用の視点

## 森 裕一 岡山理科大学 経営学部



- 兵庫県生まれ。
- 大阪教育大学、神戸大学大学院修士課程、岡山大学大学院博士課程修了。博士（学術）。
- 中学校、短期大学の教員を経て、現職。授業では、データサイエンス系の科目を担当。
- 専門は計算機統計学。
- 国内外のジャーナルの編集委員や学会の理事、国際計算機統計協会アジア地区会長（～2021年）など。
- 官公庁や民間のお手伝い、統計相談、統計教育、「統計検定」にもかかわる。

2021/07/27 おかやまデータ活用人材育成講座 WEBセミナー



# 目次

- DXとは…
- DXとデータ活用…
- データ活用ベーシック…
- データ活用に向けて…



# DXとは…

# Digital Transformation (DX) とは

## 経済産業省 (DX推進ガイドライン、2018/12)

- 企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること

AI、IoT、ICT、クラウドサービス など

## IPA (情報処理推進機構) (DX推進人材の機能と役割のあり方に関する調査、2019)

- デジタル技術の活用によって企業のビジネスモデルを変革し、新たなデジタル時代にも十分に勝ち残れるように自社の競争力を高めていくこと
- AIやIoTなどの先端的なデジタル技術の活用を通じて、デジタル化が進む高度な将来市場においても新たな付加価値を生み出せるよう従来のビジネスや組織を変革すること

企業が…  
ビジネスで…

新たな価値  
企業変革  
競争力

将来のデジタル市場  
で勝ち残る

## エリック・ストルターマン (ウメオ大学、スウェーデン、Information Technology

and the Good Life、*IFIP International Federation for Information Processing book series*, 2004)

- ITの浸透が人々の生活をあらゆる面でより良い方向に変化させること  
The digital transformation can be understood as the changes that the digital technology causes or influences in all aspects of human life.

よりよい生活  
情報システム

# DX 2025年の壁

**経済産業省** (DX推進ガイドライン、2018/12)

企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること

## 2025年の崖

DXレポート～ITシステム「2025年の崖」の克服とDXの本格的な展開～、経済産業省、2018/09

市場で勝ち抜くためには**DXの推進**が必要不可欠

複雑化・老朽化・ブラックボックス化した**既存システムの残存**、**DX人材不足**

**データを活用しきれず、DXを実現できない**ため、

市場の変化に対応して、ビジネスモデルを柔軟・迅速に変更できない

**現場の反対**

- ユーザ : ・ 爆発的に増加するデータを活用しきれず、デジタル競争の敗者に  
 ・ 多くの技術的負債を抱え、業務基盤そのものの維持・継承が困難に  
 ・ サイバーセキュリティや事故・災害によるシステムトラブルやデータ滅失・流出等のリスク増
- ベンダー : ・ 技術的負債の保守・運用にリソースがとられ、最先端のデジタル技術を担う人材を確保できず  
 ・ レガシーシステムサポートに伴う人月商売の受託型業務から脱却できない  
 ・ クラウドベースのサービス開発・提供という世界の主戦場を攻めあぐねる状態に

**デジタル競争の敗者、国際競争への遅れ**

2025年以降、**最大年12兆円の経済損失**の可能性

# DX推進のシフトチェンジ

経済産業省 (DX推進ガイドライン、2018/12)

企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立すること

2025年の崖

DXレポート～ITシステム「2025年の崖」の克服とDXの本格的な展開～、経済産業省、2018/09

レガシーシステムの残存、DX人材の不足によりDXが実現できない  
市場の変化に対応したビジネス・モデル構築に対応できない = 市場での地位の喪失  
デジタル競争の敗者、国際競争への遅れ ⇒ 2025年以降、年12兆円の経済損失

誤解 「DXとはレガシーシステムの刷新」  
「現時点で競争優位性が確保できていればこれ以上のDXは不要」

→ 本来のDXが進んでいない

95%が未着手か一部部門での実施 (←意識の高い企業の調査結果)

DXレポート2 中間取りまとめ (経済産業省、2020/12)

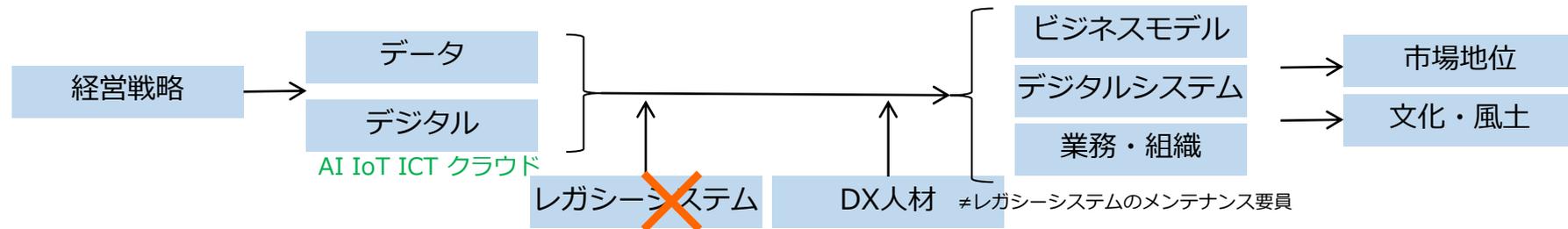
- 事業変革の**環境整備** (共通理解の形成、社内推進体制の整備など)
- **競合他社との協調領域**の形成
- 変革を対等な立場で伴走できる企業との**パートナーシップ**の構築
- 変革を遂行する**人材の確保**

論点のアップデート  
ITシステムの問題

組織・企業文化刷新の問題

# DX推進

世界中で新たなデジタル技術を利用したこれまでにないビジネスモデルが誕生・普及



**DX ≠ 単なるデジタル化**

- × オンラインツール導入
- × Web会議導入

**はばむもの**

- 日本の風習（対面重視、専門職化、法律）
- カスタマイズ型自社システム
- 現場の反対

**経営陣の考え方**

ITシステムの刷新

↓  
組織・企業文化の刷新へ

## DX：デジタルを使ってビジネスモデルに変革を起こす

企業、顧客、社会全体の生活スタイルの変革  
製品やサービスのデジタル化

デジタイゼーション（digitization）・・・業務のデジタル化  
 ↓  
 デジタライゼーション（digitalization）・・・業務プロセスのデジタル化  
 ↓  
 DX = 業務全体のデジタル化



# DXとデータ活用…

# DX推進への視点

## 組織・企業文化刷新の問題

企業文化の刷新においては、経営トップのコミットメントも必要だが・・・

- 「デジタルとかDXという言葉なるべく使わないようにしてきた。その代わりに実務アクションに一番近い“データ活用”という言葉を使い、データ活用プロジェクトというような形でプロジェクト名に冠をつけて社内浸透をさせてきた。」
- 「DXという分野で起きている一番の変化は、データ活用人材を増やそう、人材に投資しようという会社がものすごい勢いで増えている。」
- 「社員一人一人のデジタルやデータ活用に対するリテラシーや知識、ノウハウの習得が必要な時期に差し掛かっている。」
- 「経営者の隣にデータサイエンスを。」

# DX推進…データ活用はなくてはならない

## DXの話題

- 必ず データ活用
- データ分析
- データドリブン
- データサイエンス
- ⋮
- ⋮
- AI IoT クラウド
- ⋮
- ⋮

→ DXを支える**本質**は**データ活用**

↑  
DX推進の背景

ただし

データ活用さえすれば **= DX**  
は勘違い

データ活用**から始めて** ⇒ **DX**  
の流れが大切

→ **データ**

## DX推進

**データ活用** 導入

- データ収集
- データ分析
- データ蓄積
- ⋮

← 当然、そのための  
マシン・インタフェース・クラウド…

**データ**

⋮  
使いたい形で  
スピード・アジリティ  
全社（全部門）最適性

**デジタル技術**  
AI IoT クラウド…

**ITシステム** 導入

- リニューアル
- 新たな仕組み
- ⋮

← そこに、必ず  
データ活用の仕組みを入れる

## 事例から見る データ ⇔ デジタル

## オンラインショップ

- 購入行動の分析、消費者の志向
  - レコメンデーション、協調フィルタリング

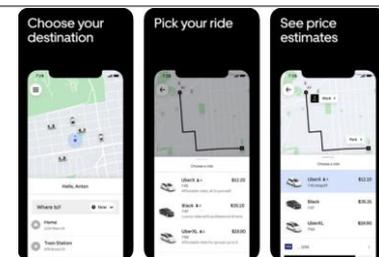
- オンライン販売インターフェース
  - 在庫・配送管理



## 配車サービス (デリバリーサービス)

- 配車・リクエスト行動の分析
  - 履歴分析、最適配置、経路選択

- 配車リクエストインターフェース
  - 省力化、最適配車、決済、雇用形態



## コーヒーマシン

- ユーザーの分析 (嗜好、頻度)、消費動向の把握
  - ポイント制度、商品管理、安否確認にも

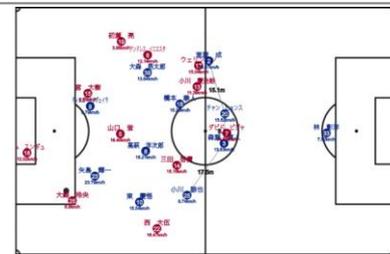
- コーヒーマシンの貸与
  - 専用アプリ



# 典型的なDX事例

## スポーツ・アナリティクス

- データ収集、アナリティクス手法、AI未来予測、戦術
  - スタッツ、リアルタイムデータ、動きの可視化、組織・選手の特徴
  - カメラ、GPS、分析ソフト、データ蓄積
    - 画像処理、位置情報、動きのトレース



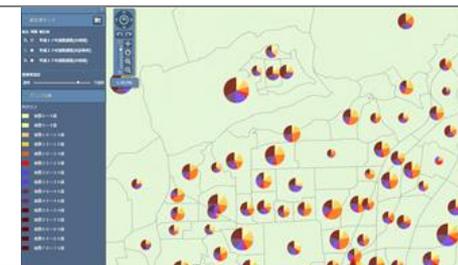
## 学生ポートフォリオ—学生情報の見える化—

- 学生自身の目標管理、教員のアドバイジング
  - 目標、適性、成績、マッチング、大学IR
  - スマホアプリ、データの一元管理
    - 学内各DB（学籍、成績）とリンク



## (書店の来店)

- 来店地域の情報把握（世帯構成、年齢構成…）⇒ 品揃えに反映
  - × 独自の地域調査 ○ オープンデータ（e-stat、RESAS）
  - システム開発不要 ⇒ 常にチェック
    - 会議で共有





# データ活用ベーシック…

# 最近のデータサイエンスの学習から

## 教養としてのデータサイエンス

(北川・竹村編、講談社、2021)

### 第2章 [基礎] データリテラシー

#### 2.1 データを読む

1. データの種類
2. データの分布と代表値
3. 代表値の性質の違い
4. データのばらつき
5. 観測データに含まれる誤差の扱い
6. 打ち切りや脱落を含むデータ…
7. 相関と因果性
8. 母集団と標本抽出
9. クロス集計表、相関係数行列…
10. 統計情報の正しい理解

## データサイエンス入門

(竹村 他、学術図書出版、2019)

### 第2章 データ分析の基礎

1. ヒストグラム・箱ひげ図・平…
2. 散布図と相関係数
3. 回帰直線
4. データ分析で注意すべき点

### 第3章 データサイエンスの手法

1. クロス集計
2. 回帰分析
3. ベイズ推論
4. アソシエーション分析
5. クラスタリング
6. 決定木
7. ニューラルネットワーク
8. 機械学習とAI

## 統計検定「データサイエンス基礎」

(2021/07 ~)

- データベース・データマネジメント  
(並べ替え、抽出、乱数、変換など)
- データの可視化  
(グラフ作成)
- 質的データの分析  
(パレート分析、連関分析、クロス集計など)
- 量的データの分析  
(ヒストグラム、基本統計量、箱ひげ図、回帰分析など)
- 確率による意思決定  
(確率計算、分布、信頼区間、推定、検定など)
- 時系列データの分析  
(移動平均、伸び率、季節調整など)
- テキストマイニング

# ①ばらつき

測定・観察されたデータには、必ず**ばらつき**がある

## ■内閣支持率30%

ま、それぐらいだろう。  
新聞社によっても違うし。  
**全員調べてない**だから当然**誤差**があるよね。

32%から30%で「下がった、下がった」は大げさ。  
70%から30%・・・「うーん、これは落ちた」な。

## ■おいしい手作りアンパン 1つ100g

1つ160円だけど、大きさに**ばらつき**があるんだよね。  
手作りだから仕方ないか。おっきめのを買おっと。

できるだけ同じ重さになるように機械で作ってるけど、  
完全に同じにはできないよね。

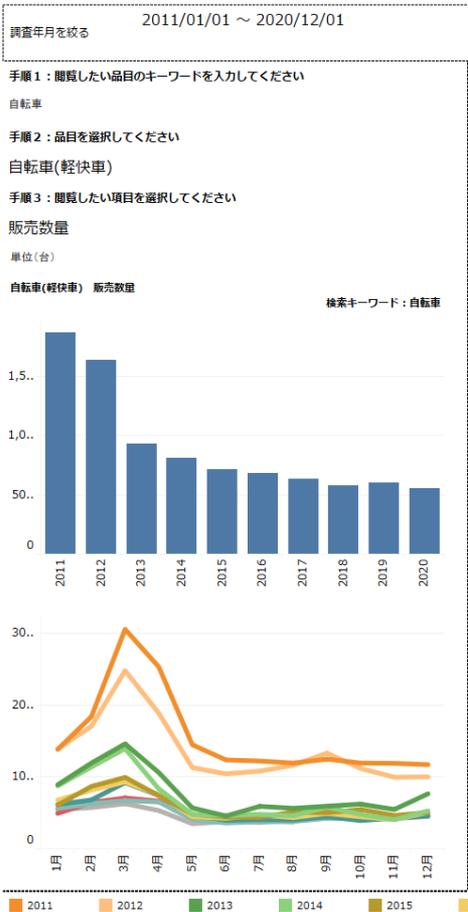
## ■視聴率30%

ビデリサーチ社によると、岡山香川地区の調査世帯数は200。  
これは、**最大7%の誤差**を認めていることになる。  
具体的に見てみると、視聴率20%の誤差は5.7%、視聴率30%の誤差は6.5%  
なので、それぞれの真の視聴率は、14.3%~25.7%と23.5%~36.5%の間。  
えー、じゃ、ひょっとして、20%の番組の方が30%の番組より  
多くの世帯が見ているという**逆転もある**わけだ・・・。

うしろに**母集団**！（偶然の誤差の存在）**ばらつき**を意識 ⇒ 「**点**」で勝負 → 「**範囲**」で比較

# ② 可視化 (内訳 | 比較 | 推移 | 分布 | 関係 | 層別 | インタラクティブ)

自転車（軽快車）の販売数量



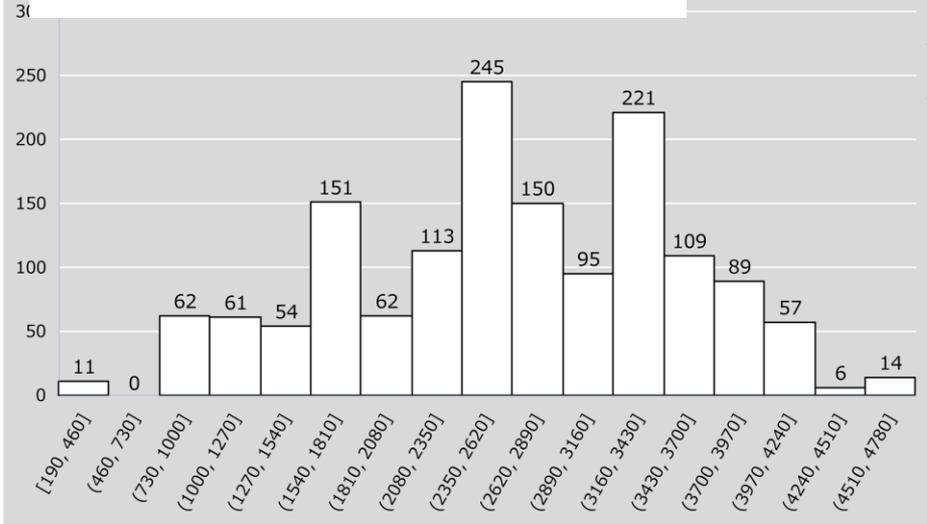
自転車（電動）の販売数量



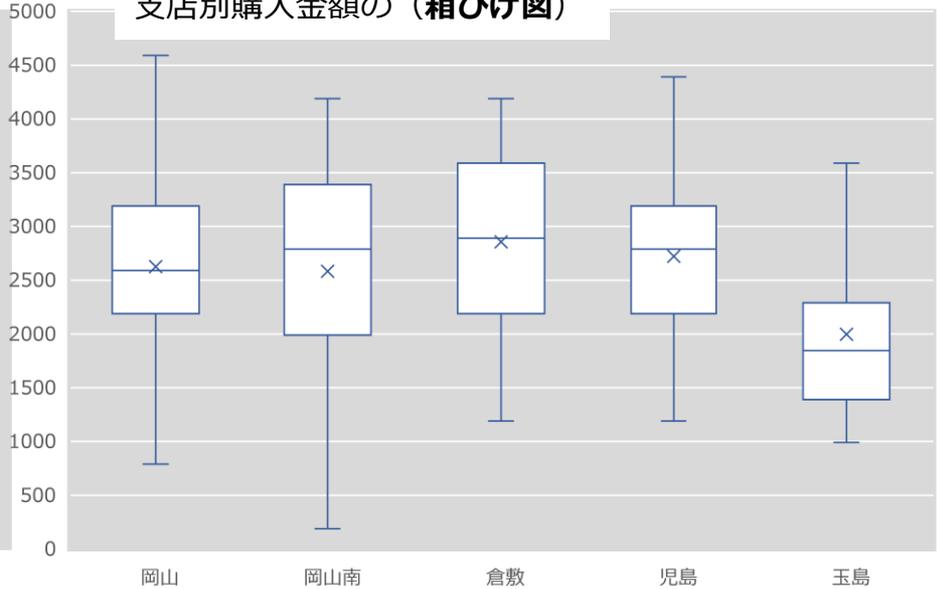
# ②可視化 (内訳 | 比較 | 推移 | 分布 | 関係 | 層別 | インタラクティブ)

回答者ID	調査支店	性別	購入金額	平均滞在時間	来店回数	満足度
11001	岡山	男	2990	15	6	2
.....	.....	.....	.....	.....	6	2

購入金額の価格帯別人数 (ヒストグラム)



支店別購入金額の (箱ひげ図)



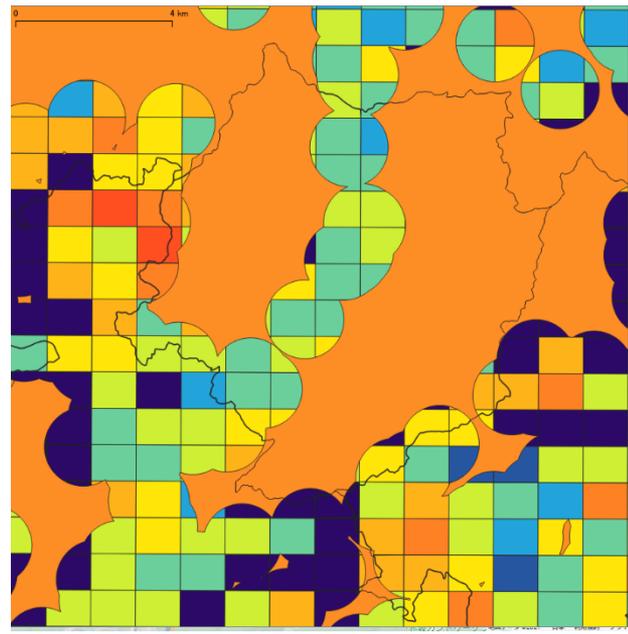
15175	玉島	女	2090	20	3	2
15176	玉島	女	2290	37	1	1
15177	玉島	女	1200	36	2	1
15178	玉島	女	1680	21	1	3
15179	玉島	女	1790	31	2	1
15180	玉島	女	2130	25	3	1

# ② 可視化 (内訳 | 比較 | 推移 | 分布 | 関係 | 層別 | インタラクティブ)

## 地図 (店舗図や設備構成図も) との融合



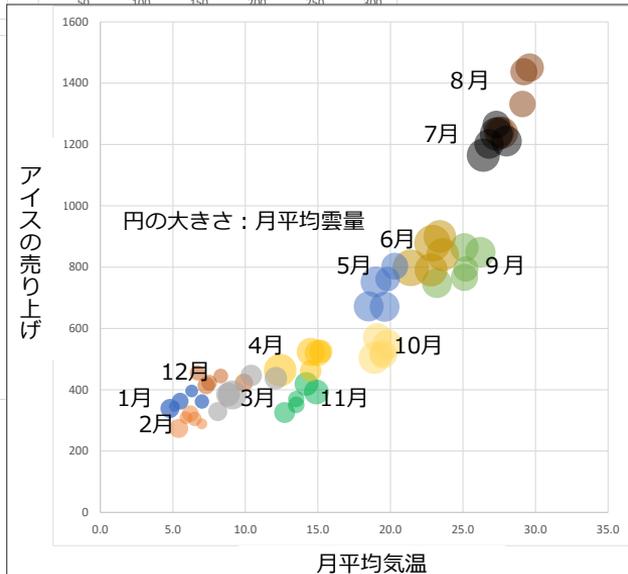
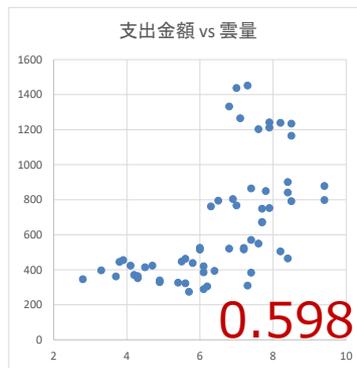
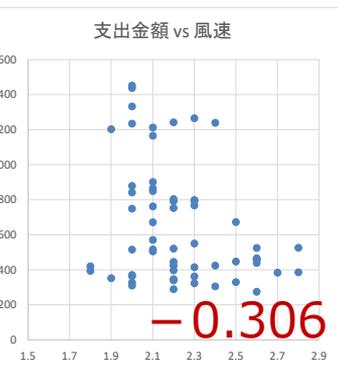
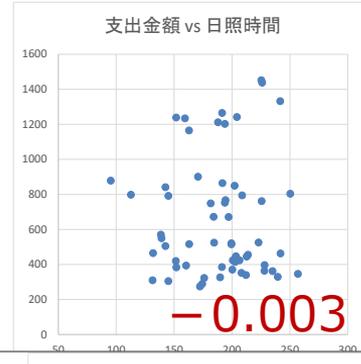
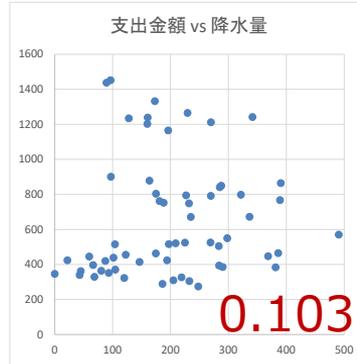
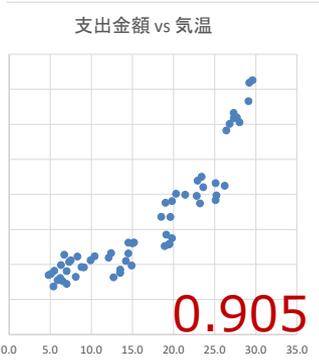
久米南町の買物弱者地域の推定



# ② 可視化 (内訳 | 比較 | 推移 | 分布 | **関係** | 層別 | インタラクティブ)

## アイスクリーム支出金額と天候の関係

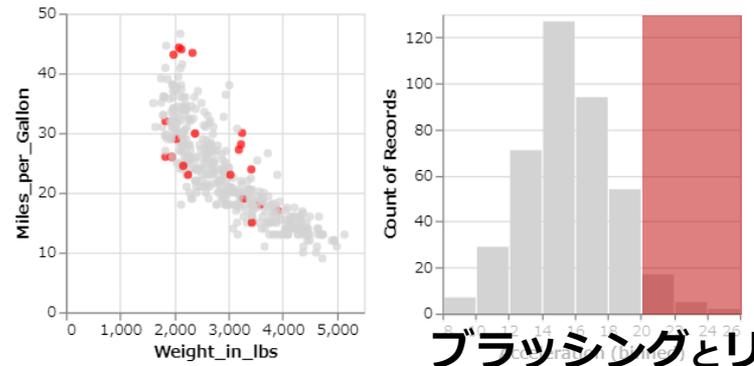
年	月	月平均気温 (°C)	降水量の合計 (mm)	日照時間 (時間)	平均風速 (m/s)	平均曇量 (10分比)	アイスクリーム支出金額 (円)
2010	1	7.0	45	235	2.3	3.7	362
2010	2	6.5	232.5	145.1	2.4	6.2	305
2010	3	9.1	381.5	152	2.7	7.4	383
2010	4	12.4	386	131.9	2.6	8.4	464
2010	5	19.0	188	193.9	2.2	7.9	752
2010	6	23.6	285	142.6	2	8.4	841
2010	7	28.0	270	188	2.1	7.9	1211
2010	8	29.6	96.5	225.3	2	7.3	1451
2010	9	25.1	390.5	191.7	2.1	7.4	864
2010	10	18.9	283.5	142.5	2.1	8.2	504
2010	11	13.5	93.5	208.2	1.9	4.3	351
2010	12	9.9	194	200.7	2.4	4.7	423
2011	1	5.1	0	256.9	2.2	2.8	346
2011	2	7.0	186	174.3	2.2	6.1	289
2011	3	8.1	68.5	239.5	2.5	4.9	329
2011	4	14.5	174.5	241.9	2.6	5.6	462
2011	5	18.5	336.5	184.1	2.5	7.7	672
2011	6	22.8	269.5	145	2.2	8.5	791
2011	7	27.3	229.5	191.5	2.3	7.1	1265
2011	8	27.5	341.5	204.2	2.2	7.9	1241
2011	9	25.1	389	194.4	2.3	7	767
2011	10	19.5	197	163	2.1	7.2	516
2011	11	14.9	284	160.4	1.8	6.4	393
2011	12	7.5	22	206.5	2.2	4.1	423
2012	1	4.8	43	212.1	2.2	4.9	339
2012	2	5.4	248	172.4	2.6	5.7	274
2012	3	8.8	290	191.3	2.8	6.1	385
2012	4	14.5	225	184.6	2.6	7.2	524
2012	5	19.6	235	197.1	2.1	7.7	671
2012	6	21.4	321.5	112.7	2.3	9.4	798
2012	7	26.4	196	162.9	2.1	8.5	1165



散布図 相関係数

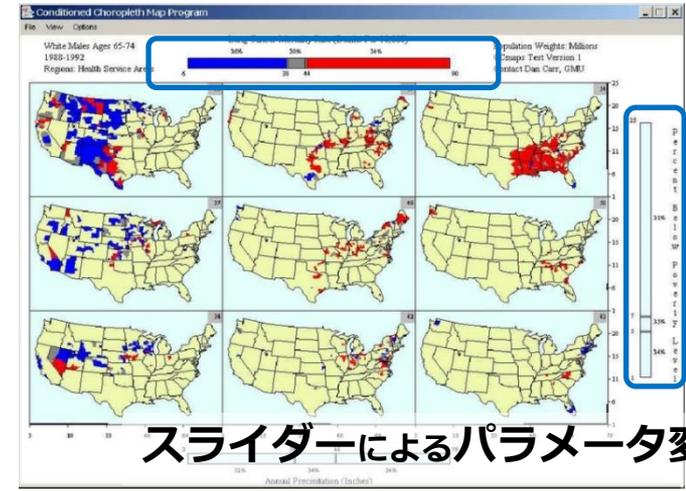
# ②可視化 (内訳 | 比較 | 推移 | 分布 | 関係 | 層別 | インタラクティブ)

Brushing and linking



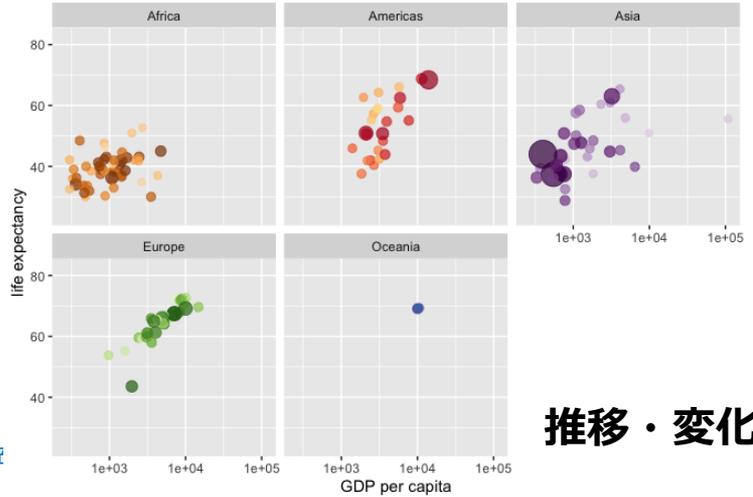
ブラッシングとリンクング

CCMap (コロプレスマップ)



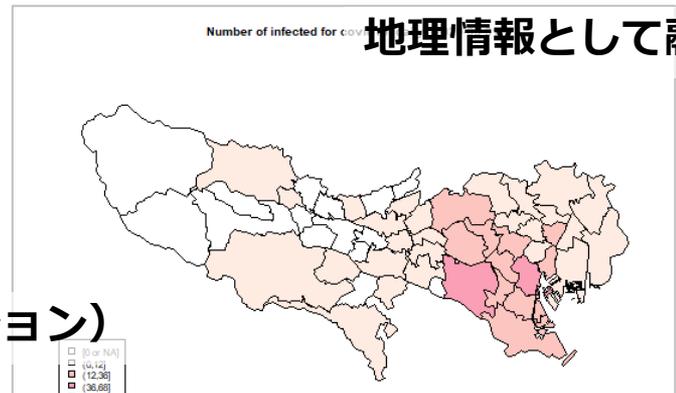
スライダーによるパラメータ変更

Year: 1952

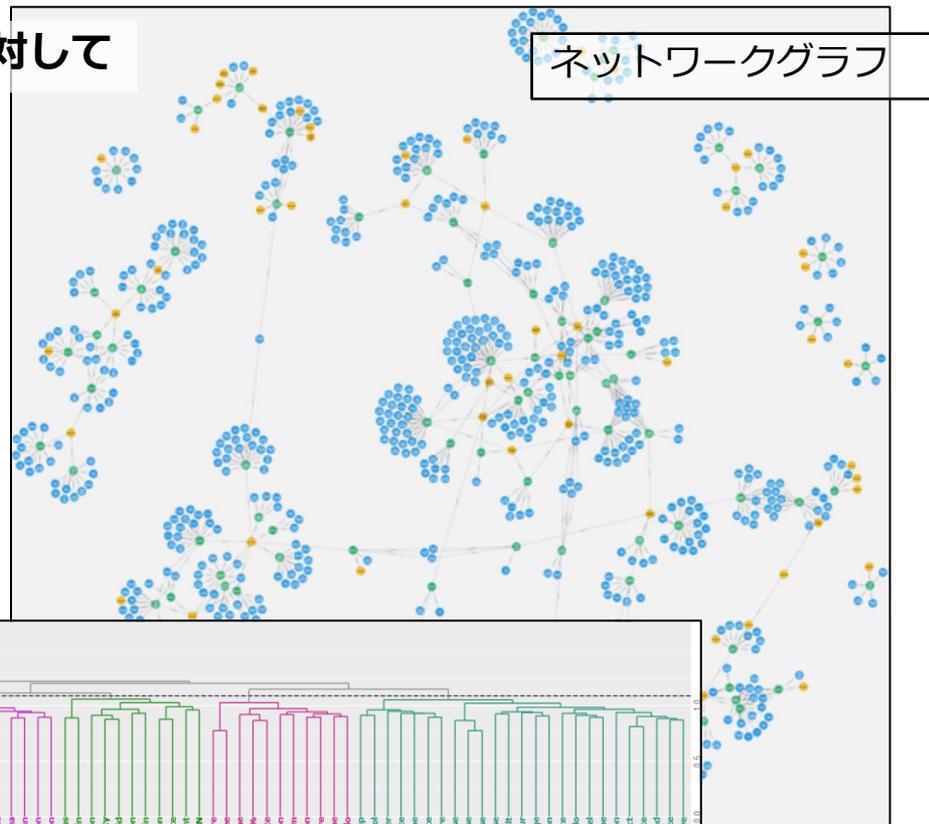
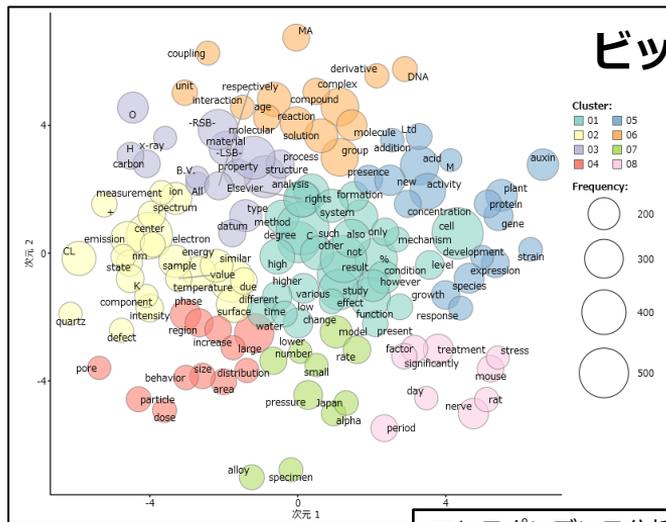


推移・変化 (アニメーション)

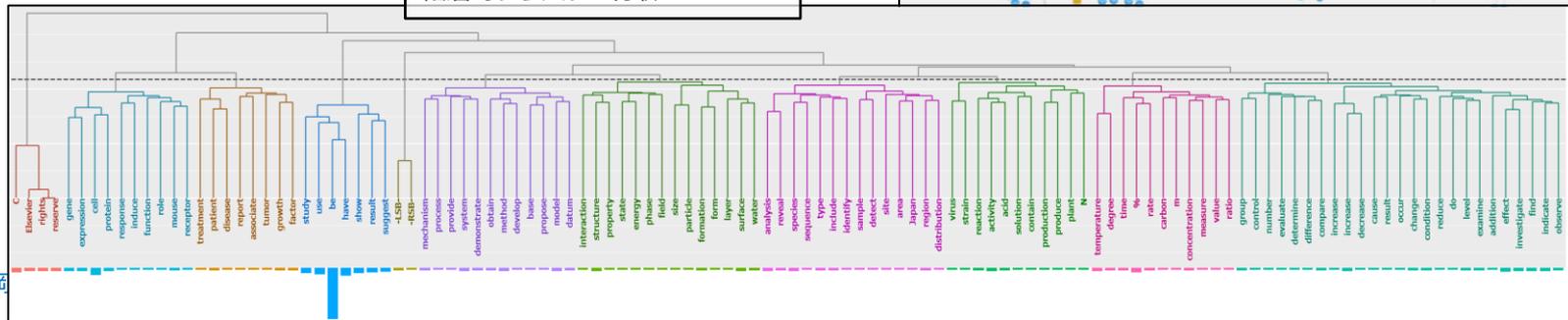
地理情報として融合



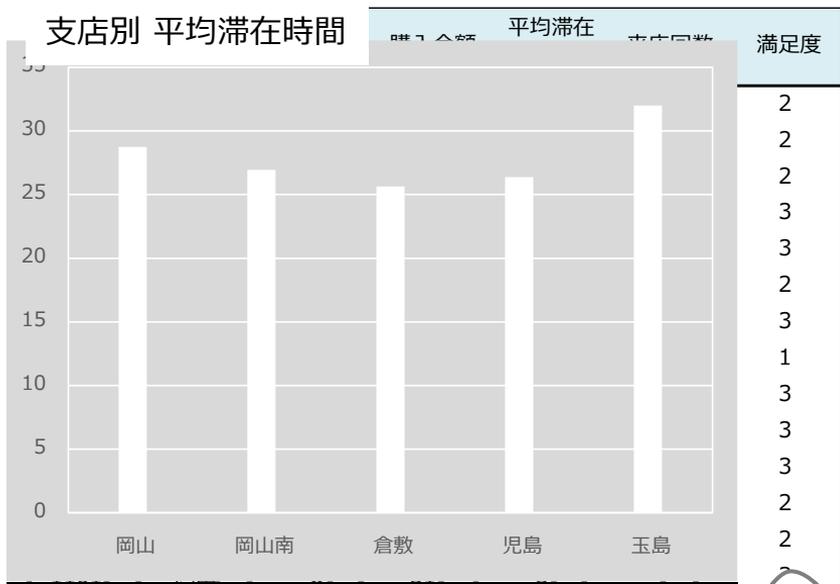
# ②可視化 (内訳 | 比較 | 推移 | 分布 | 関係 | 層別 | インタラクティブ)



コレスポンデンス分析  
テキストマイニング  
階層的クラスター分析



# ③層別・層化

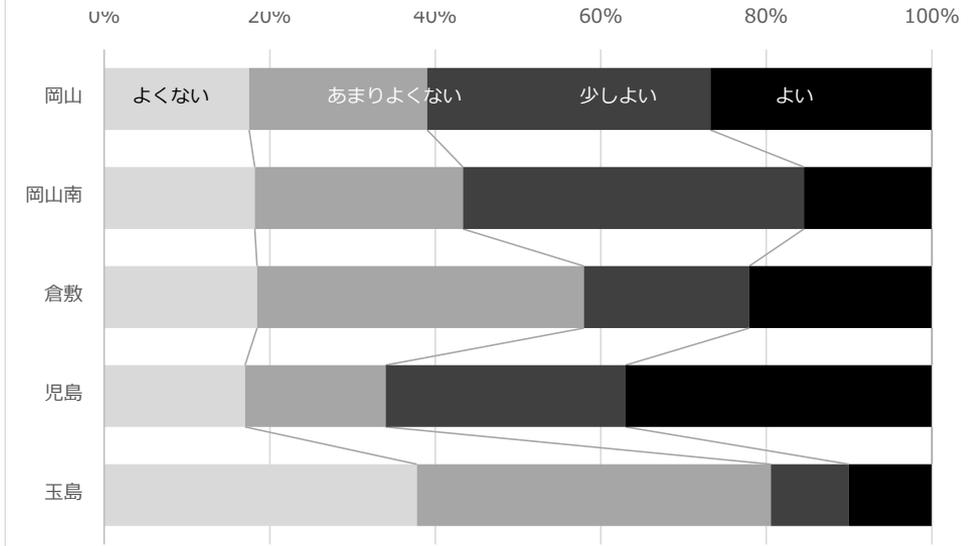


支店	平均滞在時間		計
	(男)	(女)	
岡山	27.5	29.9	28.7
岡山南	26.9	26.9	26.9
倉敷	24.9	26.5	25.6
児島	26.8	25.7	26.4
玉島	32.8	31.7	32.0
計	26.9	28.3	27.6

満足度（性別とのクロス集計）

性別	よくない	あまりよくない	少しよい	よい	計
男	9	161	259	301	730
女	295	273	167	35	770
計	304	434	426	336	1500

満足度（支店とのクロス集計）



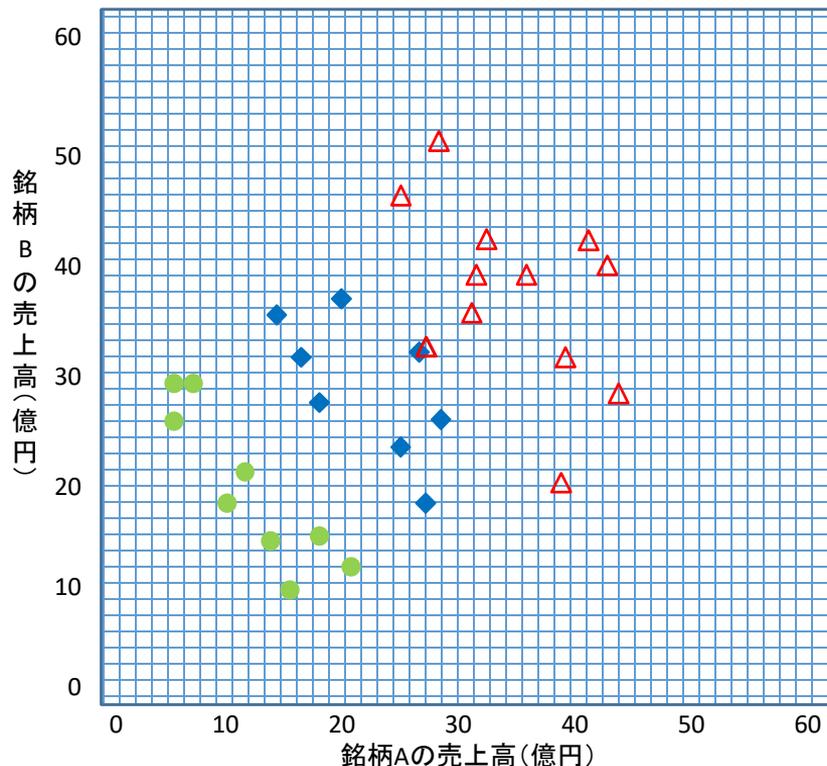
# ③層別・層化

2つの飲料銘柄の売上高（関西の29都市）

銘柄Aが売れるほど  
銘柄Bが売れている。

||  
銘柄Aと銘柄Bには  
正の相関??

現場では・・・  
両方同時に売れている  
感覚はない・・・！！



銘柄Aが売れると  
銘柄Bが売れない。

||  
銘柄Aと銘柄Bは競合！  
(一方を買えば他方は不要)

||  
銘柄Aと銘柄Bには負の相関

# 最近のデータサイエンスの学習から

## 教養としてのデータサイエンス

(北川・竹村編、講談社、2021)

### 第2章 [基礎] データリテラシー

#### 2.1 データを読む

1. データの種類
2. データの分布と代表値
3. 代表値の性質の違い
4. データのばらつき
5. 観測データに含まれる誤差の扱い
6. 打ち切りや脱落を含むデータ…
7. 相関と因果性
8. 母集団と標本抽出
9. クロス集計表、相関係数行列…
10. 統計情報の正しい理解

## データサイエンス入門

(竹村 他、学術図書出版、2019)

### 第2章 データ分析の基礎

1. ヒストグラム・箱ひげ図・平…
2. 散布図と相関係数
3. 回帰直線
4. データ分析で注意すべき点

### 第3章 データサイエンスの手法

1. クロス集計
2. 回帰分析
3. ベイズ推論
4. アソシエーション分析
5. クラスタリング
6. 決定木
7. ニューラルネットワーク
8. 機械学習とAI

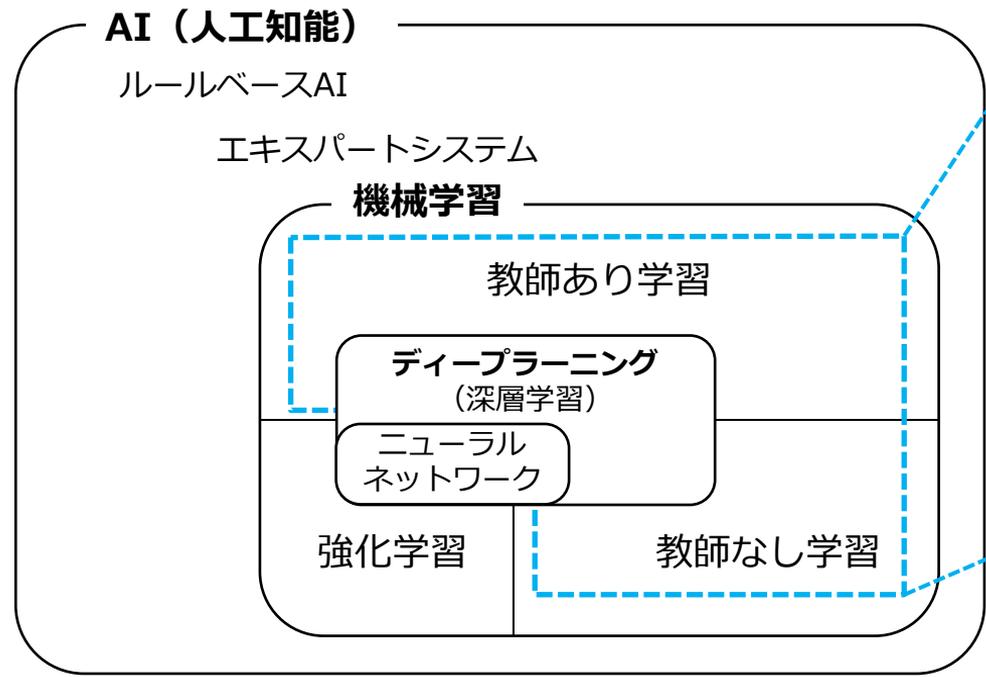
## 統計検定「データサイエンス基礎」

(2021/07 ~)

- データベース・データマネジメント  
(並べ替え、抽出、乱数、変換など)
- データの可視化  
(グラフ作成)
- 質的データの分析  
(パレート分析、連関分析、クロス集計など)
- 量的データの分析  
(ヒストグラム、基本統計量、箱ひげ図、回帰分析など)
- 確率による意思決定  
(確率計算、分布、信頼区間、推定、検定など)
- 時系列データの分析  
(移動平均、伸び率、季節調整など)
- テキストマイニング

# 機械学習とデータ分析

## AI、機械学習、ディープラーニング



### 機械学習モデル

実証済みの理論がなくても  
モデルを構築  
エラーを減らす

### 統計モデル

実証済みの理論がモデルの  
背後に存在  
仮説を検証する

### 教師あり学習

<h4>回帰</h4> <ul style="list-style-type: none"> <li>回帰分析</li> <li>エラスティックネット</li> </ul>	<h4>分類</h4> <ul style="list-style-type: none"> <li>ロジスティック回帰</li> <li>サポートベクターマシン (SVM)</li> <li>決定木   ランダムフォレスト</li> </ul>
--	---

### 教師なし学習

<h4>クラスタリング</h4> <ul style="list-style-type: none"> <li>クラスター分析</li> <li>k-means法</li> </ul>	<h4>次元縮約</h4> <ul style="list-style-type: none"> <li>主成分分析</li> </ul>
--	---

### 異常検知

- 主成分分析
- K近傍法
- 1 class SVM

利用するのは統計的手法

AINOW (<https://ainow.ai/2019/11/26/180809/>) より改変



# データ活用に向けて…

# データ活用の現場で

## うまくいかない組織

### 低い

データ活用の知識がない  
組織内の理解がない

### ない (と思っている)

データ活用が求められない  
通常業務が忙しい

### ない (と思っている)

社内のデータに気づいていない  
集める努力をしない

### できない・へた

事前に方針がない  
形式だけのデータ利用  
分析手法が間違っている

### 整っていない (と思っている)

◀データ利活用の必要性▶

◀データの分析の機会▶

◀利用できるデータ▶

◀結果の読み取り▶

◀組織内の環境▶

## うまくいっている組織

### 高い

データ活用しようという姿勢  
組織内の理解がある

### 作る

データ活用しようという姿勢  
通常業務に組み込み

### ある&集める

元々ある/既存のデータを活用  
必要なものは集めるー自社関連  
ー外部データ

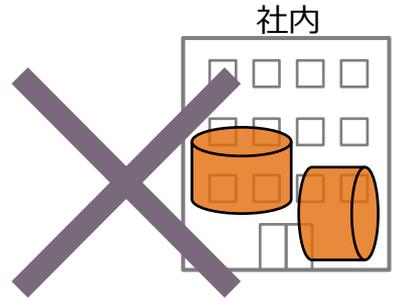
### 目的に向かって解釈

結果の活用方法が事前に明確化  
「何かを変えたい」という意気込み  
繰り返しー繰り返しー繰り返し

### 整える

ラッキー (たまたまIoTデータがあった)  
プロジェクト (トライしていく)  
そのためのシステム (新規導入)

# データ分析に対する態度



あんなデータ  
こんなデータ

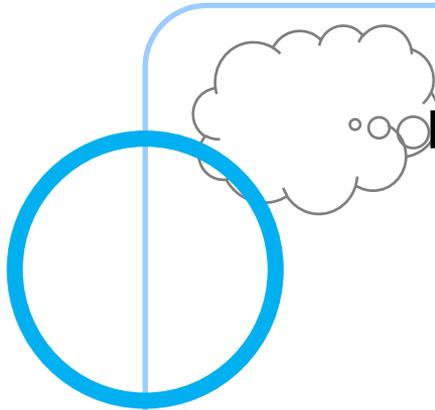
データを使う目的が**不明確**

何か触ってみる



何とか解釈してみる

何も出てこない



ゴールを予測

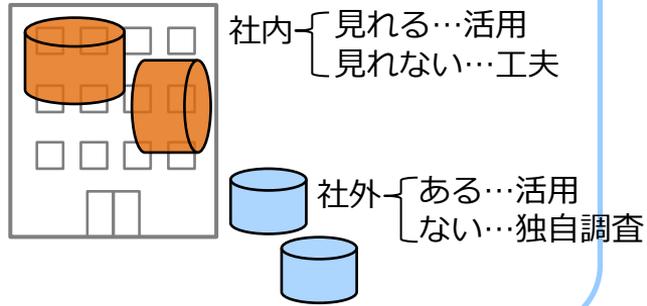


データを触る

分析…  
こうなってほしい  
こうなるにはこうであるはず  
このためにはこれがある

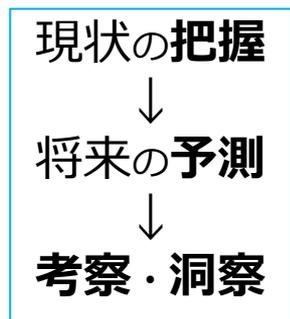
仮説

データを使う目的が**明確**



- ✓ 活用できなくてもやめない = 貴重な学び
- ✓ 分析プロセスや結果の振り返りと他部署との共有

# データとのつきあい



機械でもできる

人にしかできない →

- 基礎となる**概念**
- **観察眼**
- 構成要素で**層別** (⇒要因)

**観察** : 変化を見つけること。

**考察** : 観察から規則性や法則を導き出すこと。

**推察** : 考察によって導き出した規則性や法則の転用先を探すこと。

**洞察** : 観察・考察・推察を同時に行うこと。

= 「見えないもの」を見る / 「本当のところ」をつかむ

## 小さなデータ

仮説検定

標本の考え

全体的な傾向の把握

データ収集を慎重に

## 大きなデータ

→ 仮説検定はなくてもよい **くらい十分なデータがある**

→ **が、標本であるという意識** はもっていたい

→ 全体傾向の把握 **だけでなく、そこから はみ出る特徴**  
いろいろな層の**観察** **へ歩みを進める**

→ データがどう集められたのか **を把握することは必要**

**勝手に集まるデータ (SNSなど)** を使うときは **分析目的と方法に留意**



データ活用の文化を…  
そしてDXへ