

わたしとデータサイエンスとあなた： いままでとこれから

Me, Data Science, and You: Past and Future

19th January 2023, みんなのPython勉強会 #89

Yuta Kanzawa @yutakanzawa

Data Scientist at Zurich Insurance Company Limited, Japan Branch



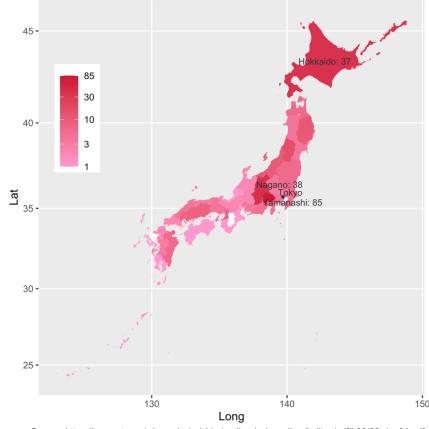
神沢雄大 Yuta Kanzawa

- データサイエンティスト@チューリッヒ保険会社
 - 日本支店
- Twitter: [@yutakanzawa](https://twitter.com/yutakanzawa)
- 好きなもの：オペラとワイン
 - ワーグナー
 - ブルゴーニュ (WSET Lv 3→?)
- 使用可能言語：7
 - 人間：日本語、英語、ドイツ語
 - コンピューター：R, Python, SAS, SQL

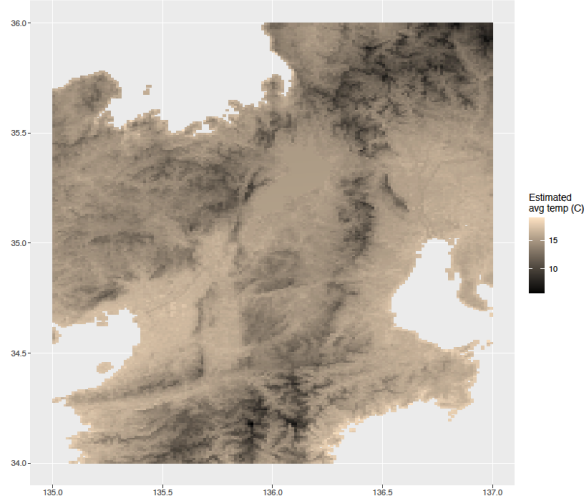


ポートフォリオ

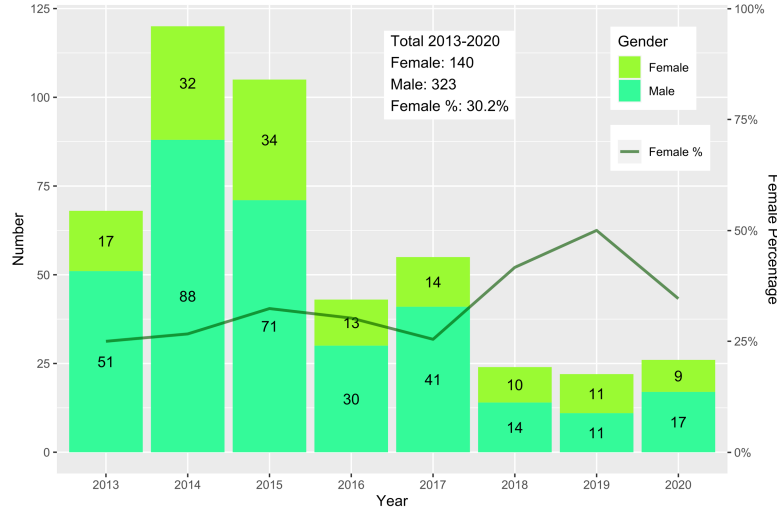
Number of Wineries in Japan in 2019, by Prefecture



Estimated Avg Temperature around Kyoto

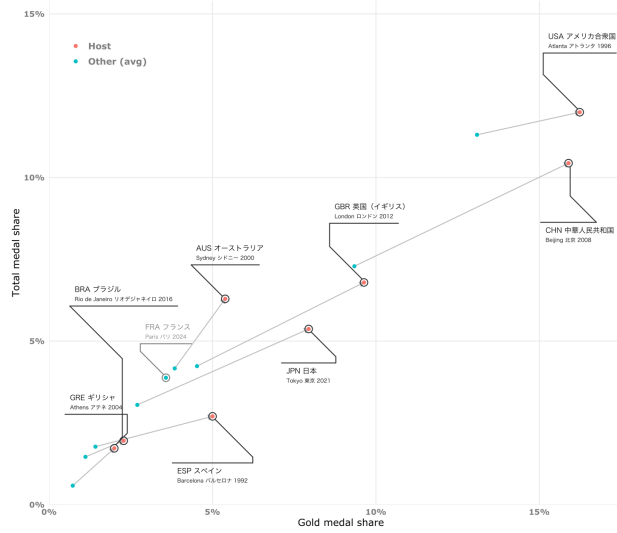


Number of Qualified JSA Sommelier Excellence and Equivalents* by Year and Gender, 2013-2020



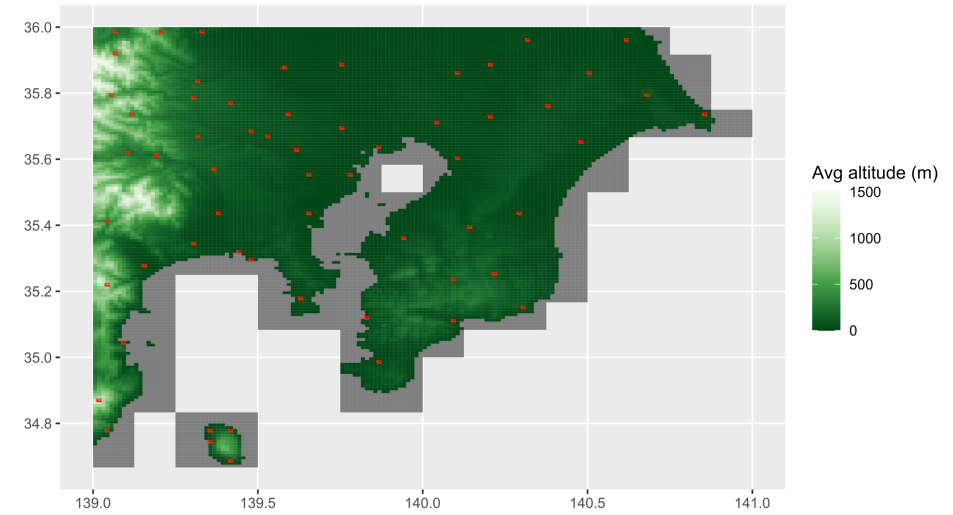
Source: Japan Sommelier Association <https://www.sommelier.jp/exam/pdf/qualifiedholders.pdf>
*Sommelier Excellence (2019-2020), Senior Sommelier (2013-2018), Senior Wine Adviser (2013-2015)

Medal shares of Olympic host countries in the past 30 years

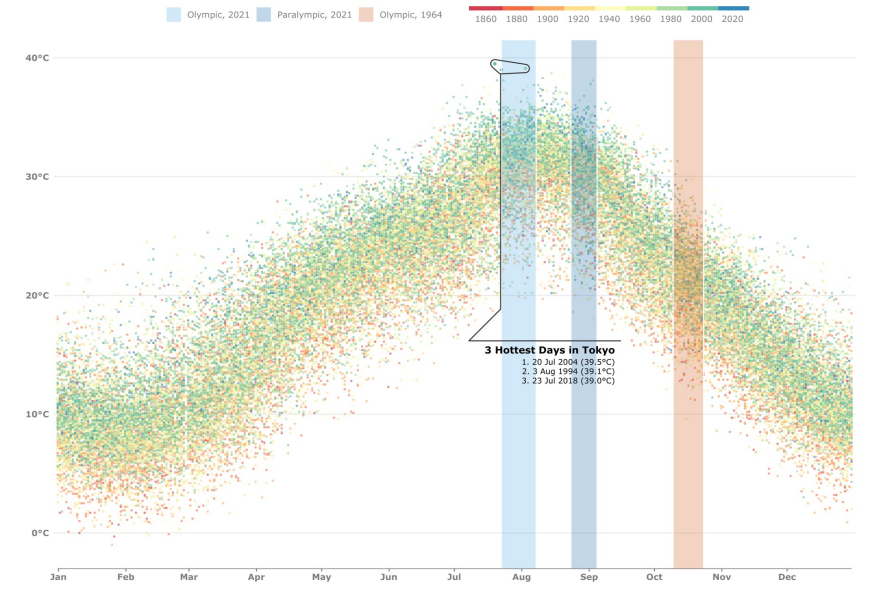


Data: International Olympic Committee via <https://olympics.com> & <https://www.wikipedia.org> - Graphic: Yuta Kanzawa

Avg Altitudes and Weather Observation Stations in Tokyo, Kanagawa, Chiba



Daily maximum temperature in Tokyo, 1875-2021



Data: Japan Meteorological Agency via <https://www.jma.go.jp> - Graphic: Yuta Kanzawa (inspired by Cédric Scherer)

(宣伝) Tokyo.R (R勉強会@東京)

明後日：2023年1月21日(土) 14-18時

<https://tokyor.connpass.com/event/271590/>



Tokyo.R @TokyoRCommunity · Jan 10

第103回の #TokyoR の参加受付を開始しました。以下のリンクよりお申し込みください。1月21日に開催予定です。応用セッションがまだ1枠ありますので、我こそはと思われる方は是非応募をお願いいたします。

多数のご応募、お待ちしております。

tokyor.connpass.com
第103回R勉強会@東京 (#TokyoR) (2023/01/21 14:00~)
第103回R勉強会@東京 (#TokyoR) を開催します！今回もオンライン開催です！* Googleフォームにて事前申告済みの発表者のみ発表者枠...



```
import qrcode  
img = qrcode.make("https://tokyor.connpass.com/event/271590/")  
img.save("TokyoR_103.png")
```

* <https://twitter.com/TokyoRCommunity/status/1612727781699510277>

アジェンダ

- 今日話すこと

- 経歴
- データサイエンティストとして
- 今後

→あるデータサイエンティストの話

- 対象（以下のいずれか）

- データサイエンティスト
- データサイエンティストの事例を知りたい人

- 今日話さないこと

- R
- Python

おことわり

- 「データサイエンス」、「データサイエンティスト」という主語の大きな話をしますが、原則としてスピーカーの実体験に基づいたものです。
- 事例を一般化するように努めていますが、全てのデータサイエンティストが当てはまる訳ではありません。
- 自分で考えて試して下さい。

TL;DR

- ビジネスサイドのデータサイエンティストとして
 - 職能に軸足
- スキル、知識：分かって使う。データサイエンス以外も大事。
 - プレゼンテーション、チームビルディング、ビジネス
- 働き方
 - 期待値マネジメント
 - 価値の提供
- これから
 - マネジメント + コードを書くスキルの維持 + 最新技術のキャッチアップ

Ch 1: わたしとデータサイエンスのいままで

Me and data science: Past

「セクシー」*と言われて丸10年

- HBRの論文（2012年10月号、邦訳2013年2月号）
 - 'Data Scientist: The **Sexiest** Job of the 21st Century'
 - <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
 - Thomas H. Davenport & D.J. Patil
 - 副題：'Meet the people who can **coax treasure out of messy, unstructured data.**'

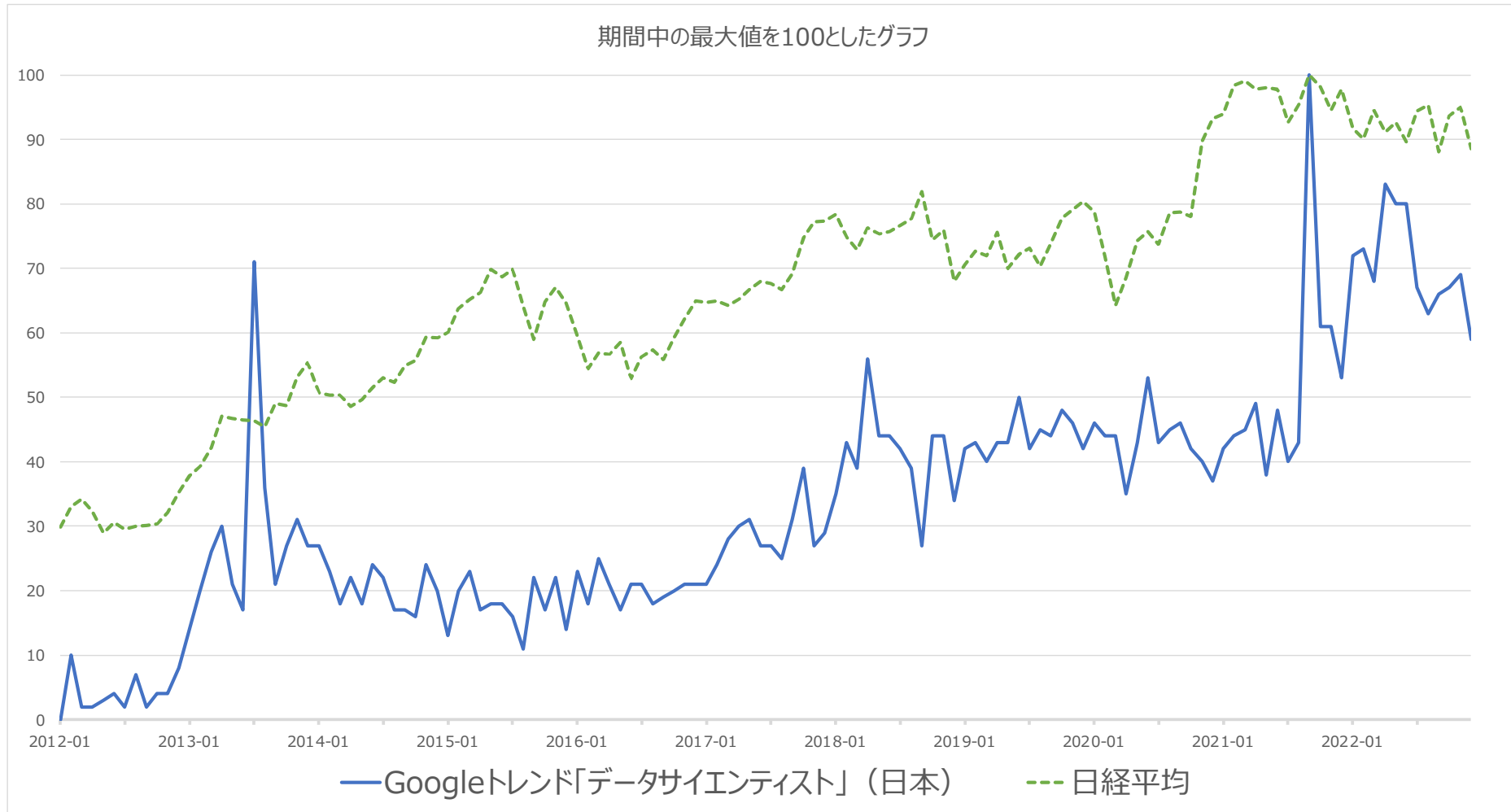


* 正確に訳すと「魅力的な」、「素敵」（邦題）。

* <https://store.hbr.org/product/harvard-business-review-october-2012/br1210>

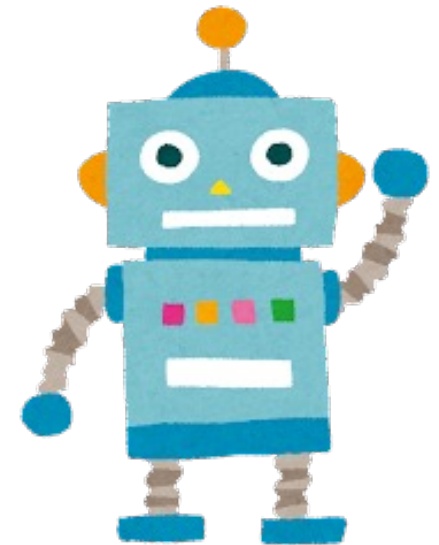
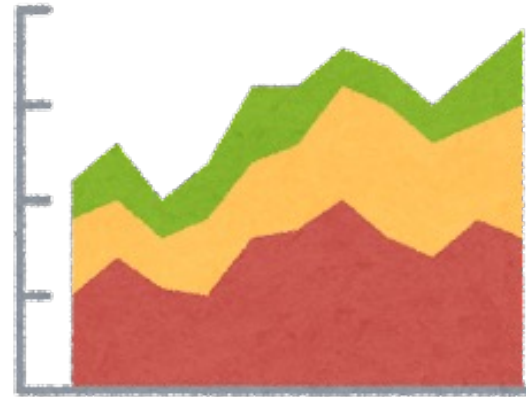
* <https://dhbr.diamond.jp/ud/backnumber/50ebdd876a8d1e3c6a000001>

「データサイエンティスト」






様々なタイトル（肩書き）

- データサイエンティスト
 - データアナリスト
 - 機械学習エンジニア
 - AIエンジニア etc
-
- 「データサイエンス」は広範な領域。
 - 機械学習/AIエンジニアは（当然ながら）エンジニアリングがメイン。



わたしのいままで：ビジネスサイドのデータサイエンティスト

会社	期間	事業内容	仕事相手	仕事内容（分野）
データ・フォアビジョン 	2012 ↓ 2016	金融業界向けソフトウェア 開発&分析ベンダー	顧客企業（銀行、カード 会社、通信事業者）	<ul style="list-style-type: none"> 信用リスク評価 マーケティング施策
ヤンセンファーマ 	2017 ↓ 2022	J&Jの製薬部門 （医療用医薬品のみ）	マーケティング、営業部門	<ul style="list-style-type: none"> マーケティング施策 営業効率化
チューリッヒ保険 	2022 ↓ 現在	日本では損害保険のみ （法的規制により別の会 社が生命保険を）	マーケティング、営業部門、 コールセンター、管理部門	<ul style="list-style-type: none"> CX向上 マーケティング施策 業務効率化

転職：理由と目的

- DFV → ヤンセン（ベンダーからユーザーへ）
 - **意思決定に近い位置**で働きたい。
 - 直接ビジネスに貢献したい。
- ヤンセン → チューリッヒ（他の業界へ）
 - **より顧客に近くダイナミックな事業環境**で働きたい。
 - 医療用医薬品は医療従事者を経て患者さんに届く（B2B2Cに近い）。
 - 競争は熾烈だが、環境変化がいつ起こるのか予め大体分かる。
 - 新薬発売
 - 特許切れ
 - 薬価改定

Ch 2: データサイエンティストとしてのわたし

Me as a data scientist

データサイエンスっぽいスキル

- プログラミング
 - PythonもRもSQLも
 - Juliaも？
 - **自分がしたい処理を書ける。**
 - 綺麗なコードで。
- データ加工
 - **tidy**なデータ（整然データ）
- データ可視化
 - matplotlib, plotly
 - ggplot2

データ可視化

データ加工

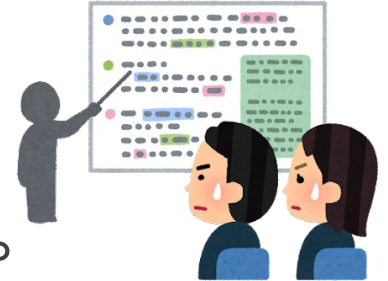
Python

R

SQL

データサイエンスっぽくない（けど大事な）スキル

- プレゼンテーション
 - 自分のしたことを**分かってもらう**。
 - 仕事相手は結果（主にプレゼンテーション）しか見ない。
- チームビルディング
 - プロジェクトや日々の業務を**円滑に進める**。
 - 仕事仲間はデータサイエンティストだけとは限らない。
- 英語
 - **会話**ができると幅が広がる。



知識

- 統計学
 - 日々の業務で数字を見る人なら基礎は理解しておきたい。
- 機械学習
 - ライブラリーを呼んで使える、では不十分。
- IT
 - ネットワーク、サーバーなど知っているといい。
- ドメイン知識
 - **自社や顧客のビジネスのことを分かっていないと、作業するだけになってしまう。**
 - 業界、商習慣、規制、マーケティング



* https://twitter.com/genbaneko_bot/status/1265783114959212544

ツール

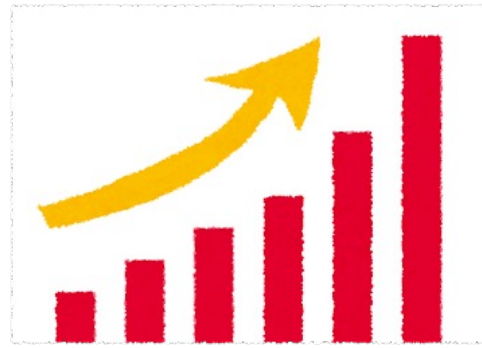
- プログラミング言語
- スプレッドシート
- ノー/ローコードツール
 - PowerBI、Tableau、Dataiku、など

→ 何を使うかよりもアウトプットが大事。

- **でも、プログラミング言語を使うと再現性と汎用性が高くなる。**



何のために働く？



- データサイエンスの楽しさ
 - データから合理的に価値を生み出す。
- お金
 - スキルと経験に見合う報酬、待遇
 - 報酬がいいからデータサイエンティストをする訳ではない。
- 職業倫理
 - 科学的（批判的）である。
 - 顧客のために。
 - 'The patients are waiting.' (Dr Paul Janssen)



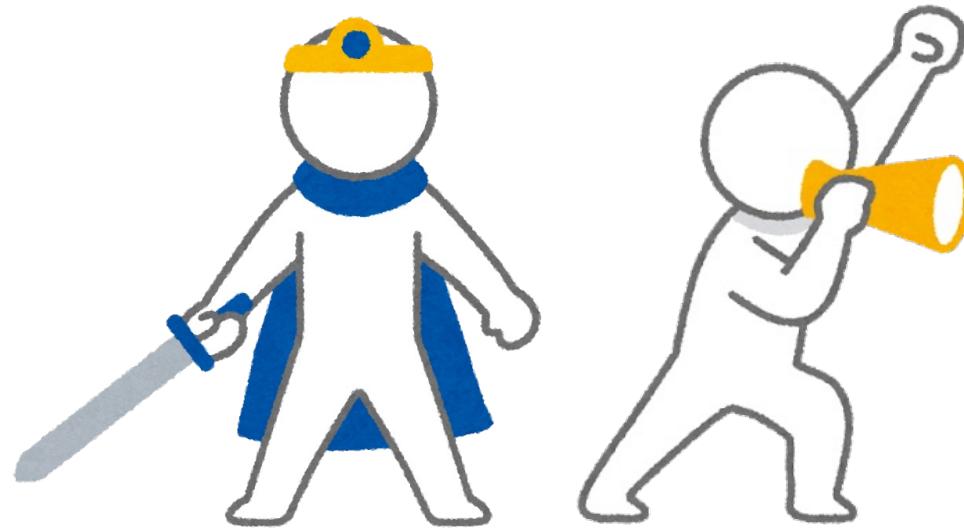
どう働く？

- 仕事相手：顧客企業やマーケティング、営業、プロダクトのチーム
 - **ビジネスの目標を共有**。→ ジブンゴトとして実現を目指す。
 - **期待値マネジメント** → データサイエンティスト≠魔法使い
 - データサイエンス関連でない人 → 丁寧に説明する。例：専門用語の回避
- 「結果を出す」
 - どのようにどんな**価値を提供**するのか？
 - アウトプット：インサイト、レポート、予測値
 - ゴール：売上達成？ KPI改善？
 - コストセンター
 - データサイエンスチーム単独では**利益を生み出さない**。



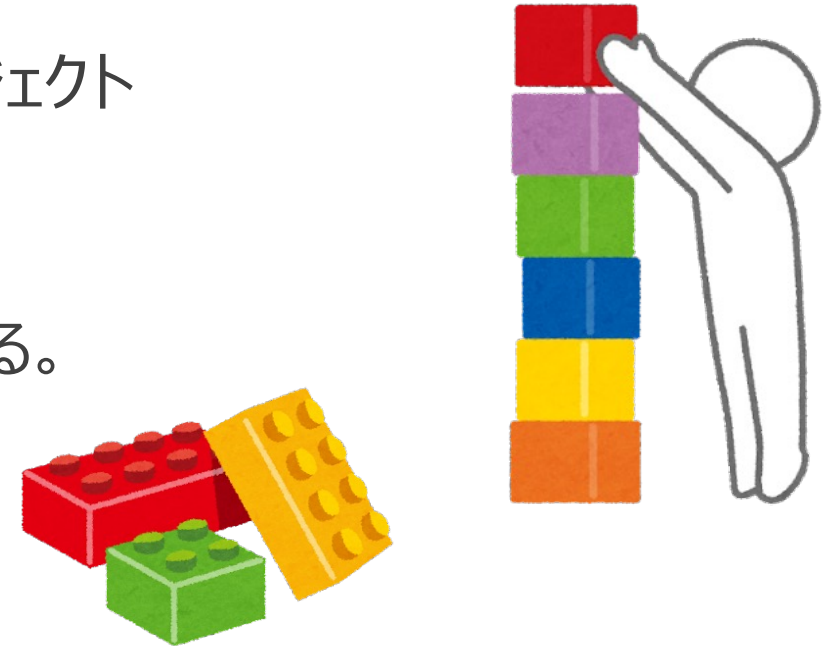
どういうポジション（ジョブ）で働く？

- スペシャリストとマネージャー
 - ピープルマネジメントをするかどうか。
 - スペシャリスト：個人の能力を最大限に発揮。
 - マネージャー：チームのレバレッジを利かせる。



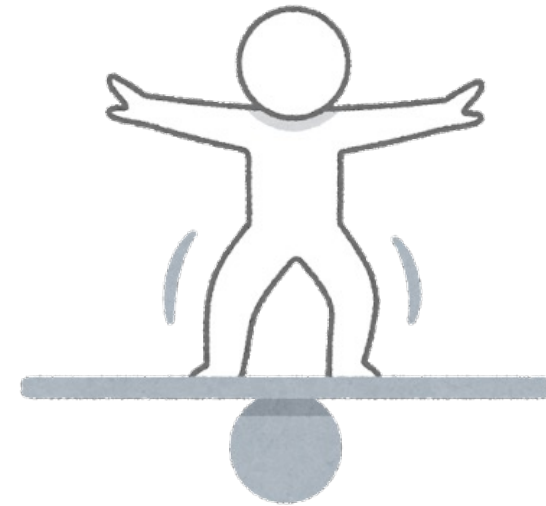
キャリア形成 (1/2)

- 経験をどう積む？
 - 中途採用で求められているのは**経験のある**データサイエンティストというジレンマ
 - 小さく始めてみる。例：日々の業務、個人プロジェクト
- ベンダー側とユーザー側
 - ベンダー：
 - 長所：**色々な顧客（業界）**のデータに携われる。
 - 短所：意思決定から遠い。
 - ユーザー：
 - 長所：**意思決定**に近い。
 - 短所：業界、データは同じ。



キャリア形成 (2/2)

- 副業としてのデータサイエンティスト（特にユーザー側にいる時）
 - 本業では経験できないことを通して**スキルと経験を広げる**。
 - 異業種、新しい種類のデータ
 - 注意
 - あくまで**本業が最優先**。「複業」ではない。
 - **競業避止義務**
- データサイエンス × 業界
 - どちらに軸足を置くか。
 - 職能としてのデータサイエンス：業界により求められる手法や経験は異なる。
 - 業界の知識、経験：同業他社では活かせても、他業界ではダメかも。



Ch 3: わたしとデータサイエンスのこれから

Me and data science: Future

データサイエンスとデータサイエンティストの今後

SWOT分析*	プラス要素	マイナス要素
内部環境	Strength (強み) <ul style="list-style-type: none">機械学習やAIの発展<ul style="list-style-type: none">精度向上複雑化活発な知見の共有 (コミュニティ)	Weakness (弱み) <ul style="list-style-type: none">偏った男女比モデルやインサイトのバイアスデータリテラシーの不足 (一部)
外部環境	Opportunity (機会) <ul style="list-style-type: none">デジタルトランスフォーメーション (DX)<ul style="list-style-type: none">機械学習やAIの活発な利用ビジネスにおけるデータドリブン化紙のデータ化人材供給	Threat (脅威) <ul style="list-style-type: none">機械学習やAIの発展<ul style="list-style-type: none">ノー/ローコードツール自動化ツールデータリテラシーの不足

* <https://ja.wikipedia.org/wiki/SWOT%E5%88%86%E6%9E%90>

セクシーだがセクシストではない。

- 男女比
 - ITエンジニア → 8:2 (2020年版情報サービス産業基本統計調査)
 - データサイエンティスト → 9:1 (データサイエンティスト協会、2021年)
- DE&I
 - モデルやインサイトが**社会的バイアスを助長**していないか？
 - 例：住宅ローン審査モデルの特徴量の検討
 - 年齢
 - 性別
 - 永住権有無



* <https://www.jisa.or.jp/Portals/0/report/basic2020.pdf>

* https://www.datascientist.or.jp/common/docs/personal_research2020.pdf

わたしのこれから（妄想）

短期（0.5-1年）

- 保険業界、ビジネスの理解
- データの把握
- 関係構築

中期（3年）

- より大きいチームでビジネスに貢献。
- プロジェクト/ピープルマネジメントの強化

長期（5-10年）

- （遅くともこの頃までには）
国外でデータサイエンスチームをリード。
- CDOやCAO

コードを書くスキルの維持 & 最新技術のキャッチアップ

まとめ

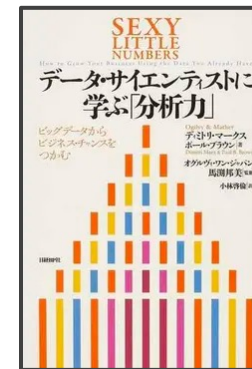
Long story short

Long story short

- ビジネスサイドのデータサイエンティストとして
 - 職能に軸足
- スキル、知識：分かって使う。データサイエンス以外も大事。
 - プレゼンテーション、チームビルディング、ビジネス
- 働き方
 - 期待値マネジメント
 - 価値の提供
- これから
 - マネジメント + コードを書くスキルの維持 + 最新技術のキャッチアップ

参考書

- 『イシューからはじめよ』（安宅、2010年）
- 『データ・サイエンティストに学ぶ「分析力」』（Maex、Brown、2013年）
- 『戦略的データサイエンス入門』（Provost、Fawcett、2014年）
- 『ビジネスデータサイエンスの教科書』（Taddy、2020年）



Enjoy!