



企業戦略と政策のための実践的データ分析: 方法論とこれからの課題

一橋大学大学院経済学研究科 原泰史 yasushi.hara@r.hit-u.ac.jp

自己紹介

- 1998-2004
 - TOYOTA NATIONAL COLLEGE OF TECHNOLOGY
- 2000
 - Exchange Student in Malaysia
- 2002-2009
 - **CLARAONLINE, INC.**
 - ICT Hosting Company, nowadays called Cloud system supplier
- 2009-2015
 - Institute of Innovation Research, HITOTSUBASHI UNIVERSITY
- 2015-2017
 - **Science for RE-Designing Science, Technology and Innovation Policy Center, National Graduate Institute for Policy Studies (GRIPS) / NISTEP / Hitotsubashi UNIVERSITY/MANAGEMENT INNOVATION CENTER**
- 2018-2019
 - EHESS Paris – CEAFJP/Michelin Research Fellow
 - OECD Expert Advisory Group: Digital Science and Innovation Policy and Governance (DSIP) and STI Policy Monitoring and Analysis (REITER) project
- 2019-
 - **TDB Center for Advanced Empirical Research on Enterprise and Economy, Faculty of Economics, Hitotsubashi University**



最初に
2005年と2015年の話

2005

shutterstock.com • 501092143

ITスタートアップで
働いていたころの話 |

“IT スタートアップ” とは言っても…

- 働き出したころ(2002年) は従業員7名
- 辞めるころ (2009年) でも従業員70名
- “レンタルサーバ”と呼ばれていた時代で、クラウドコンピューティングがまだまだ普及する前 (AWS の黎明期)
- 仮想化 (virtualization) という、サーバのハードウェアレイヤーとOSレイヤーをカーネルレベルで独立させる技術が一般化しつつ会った時代 (VMware や Parallels Container など)
- Web で注文を掛けて、FAXで受注するような時代
- 購買システムと受注システムと会計システムがバラバラだった時代

VPS(バーチャルプライベートサーバ)

- サーバサービスは当時価格競争に突入しており、3000円弱の新サービス(Demi)を投入することに
- ところが、最低価格のプロダクト(Demi)ではなく、中価格帯(Solo)が最も売れるようになる
 - でも、なぜ一番安いモデルが売れないのか説明できない
- 「相対価格」で説明して、なんとなく上司が納得する(腹落ちする)
 - Demiをベースにすると, 1.67 -> 2.34 -> 3.35
- その後最安価モデルは廃止

VPSのスペック				
名称	Demi	Solo	Doppio	Quad
初期費用	無料			
月額費用※1	2,980円 (1円)※2	4,980円 (1円)※2	6,980円 (1円)※2	9,980円 (1円)※2
年額費用※1	32,780円 (26,823円)※2	49,800円 (44,823円)※2	69,800円 (62,823円)※2	99,800円 (89,823円)※2
サポート	メールのみ	メール・電話※3	メール・電話※3	メール・電話※3
ディスク容量	1GB	1GB	2GB	4GB
月間データ転送量	25GB	50GB(増量)	100GB(増量)	200GB(増量)
バーチャルホスト※4	3(サブドメイン無制限)	無制限		
VPS設定情報	VPSサービスの設定情報ページをご覧ください			
試用	OK			
ソフトウェア	無償または有償インストール			
PHP	●	●	●	●
MySQL	○	○	●	●
SSL※5	○	○	○	○

●開通時無償インストール対応 ○オプション対応



2

0

1

5

エビデンスに基づく 科学技術イノベーション政策と「ポンチ絵」

- いまままでのキャリアで、ベンチャー(スタートアップ)とアカデミアをやったことに気づき、数年だけならばと、政策大学大学院大学の文部科学省の科学技術イノベーション政策のセンター (SciREX センター) で働くことに
- どうやら、研究内容やどういうジャーナルに載ったかとか、それがどういうインパクトファクター(IF)だったかよりも、政策立案のプロセス内で使われる ポンチ絵の一部 になっていることが重要らしい。
- ポンチ絵の要件
 - 数十秒で読んで理解できること
 - しかも、内容が過不足なく明記されていること
 - 色がいっぱい使われていること
 - 何よりも予算が取れること

「ポンチ絵」



自分もこんな感じのポンチ絵をつくる

SciREXの成果が政策形成の実務に貢献した例 SciREXセンター 政策分析・影響評価領域プロジェクト STI政策の定量的評価を行うための経済モデル間の比較研究

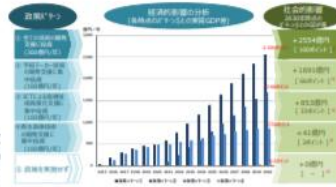
目的: STI政策を定量的に評価できる経済モデルの比較検討を行う

多部門経済一般均衡的相互依存モデル

- 利点
- 産業ごとの経済厚生の変化を描写できる
 - 政策実施による影響を把握しやすい
 - 糖尿病およびIoTモデルをこれまでに開発

モデルの構造

- 政策立案および実施時に想定されるシミュレーションを可能とするため、各産業の技術的な特性を反映させている。これにより、将来に渡る産業ごとのTFP成長率、労働人口動態(年齢別、性別)、価格変化などを年ごとに把握することが可能。

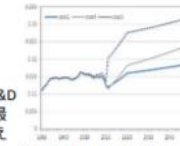


糖尿病モデルの分析イメージ
(政策ごとの経済効果を測定可能)

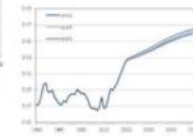
参考文献: 黒田昌裕, 池内健太, 原泰史, 土谷和之, 尾花和弥: 科学技術イノベーション政策における政策オプション作成のためのモデルの研究開発, GRIPS SciREX センターワーキングペーパー, 2015.09 (近刊)

R&D動学一般均衡 (DGE) モデル

- 利点
- 経済学の理論的根拠を重視した標準モデル。経済主体の将来的な行動を織り込んでいる。
 - モデルの構造
 - 今日のマクロ経済学の代表的な動学モデルに R&Dモデルを導入することで、企業や家計が長期的な最適化行動を行うなかで、政府の研究開発投資が与える影響を測定する



政策が民間研究開発投資に及ぼす影響の測定



政策が家計消費に及ぼす影響の測定

参考文献: 榎井誠: 科学技術イノベーション政策の経済成長分析・評価, JST/RISTEX 科学技術イノベーション政策のための科学研究開発プログラム 平成26年度プログラム 全体会議(発表資料), 2015.2

今後の課題

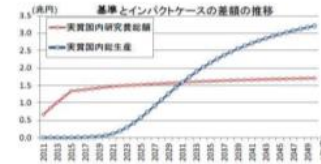
- モデルを活用した政策事例の蓄積
- モデルプラットフォームの一元化
- モデル分析の高速化・迅速化(SciREXセンター一橋大協働でスパコン京を用いた実証試験を実施中)

MaeSTIPモデル

- 利点
- マクロ経済に係る指標が全て網羅
 - 家計・企業・政府の行動間数が入り

モデルの構造

- 政府による研究開発投資が企業や大学での研究開発活動に用いられ、それらの成果が知識ストックとして蓄積されたとき、イノベーションを促進しTFP(全要素生産性)を向上させ、経済成長を牽引すると定義。これらの動態をマクロ経済指標を用いてシミュレートする。



標準モデルとイノベーションケース(科学技術関係経費が年率6%成長する場合)の潜在GDPおよびTFPの変化率

参考文献: 文部科学省 科学技術・学術政策研究所 一橋大学 イノベーション研究センター 赤池伸一 藤田健一 外木暁幸 花田真一: 科学技術イノベーション政策のマクロ経済政策体系への導入に関する調査研究, 2013.10 <http://hdl.handle.net/11035/2433>

NISTEPモデル

- 利点
- シンプルかつコンパクトなモデル設計
 - 迅速に分析を行うことが可能

モデルの構造

- 科学技術関係経費を入力データにしたマクロ経済モデルである。支出、生産、価格、雇用分配および研究開発のブロックから構成される
- GDP総額、雇用量の総額、輸出入や民間の研究開発投資の把握、研究開発投資が貿易に与える影響も分析可能

科学技術関係予算の想定シナリオごとの実質GDPへの影響

第4期及び第5期の科学技術関係予算の累計と第3期の実績値(21.7兆円)との差	2027年度までの累積実質GDPの差
①-0.3兆円(第4期分) 0.3兆円(第5期分) → +6.0兆円	+20.8兆円
②-0.3兆円(第4期分) 0.3兆円(第5期分) → +0.0兆円	+20.8兆円
③-0.3兆円(第4期分) 0.3兆円(第5期分) → +1.1兆円	+20.2兆円

※ シナリオタイプ
①予算総額25兆円, ②予算総額27兆円, ③予算総額30兆円

参考文献: 永田晃也: マクロモデルによる政府研究開発投資の経済効果の計測, DISCUSSION PAPER, 科学技術庁科学技術政策研究所, 1998.3 <http://hdl.handle.net/11035/422>

ここまでのまとめ

- (個人的には、)ITスタートアップも中央官庁も「データ分析」に対するアティチュードはそんなに違いはなかった
 - 「直感でわからないことや、うまくいかなくなったことを分析したい」ことへの高いニーズはあるし、それぞれにデータもある
 - しかし、部署ごとにデータがバラバラになっており、繋がっていない
 - 解析を行うための方法論や理論もあまりない
 - 誘導系でも構造系でも、統計的な分析モデルを構築するというよりは、棒グラフや線グラフで経過を見るだけで必要十分だった(意思決定プロセスの上位過程にまわるタイミングで、説明が出来るレベルに留めることが肝要)

2018
Paris



2019年春、パリから日本に戻ってきたら、流行っていたもの

- タピオカミルクティー



Credit: CC0

- データサイエンス



References: <https://enterprisezine.jp/bizgene/detail/4739>



2020

どうやら、この国では「データを使う」
ことが流行っているらしい

- 一橋大経済学部で「経済学のための実践的データ分析」という講義をはじめたところ、思っていたよりも沢山の学生が
- 経済学・(実証的な)経営学の学生・研究者向けに Python 3.x を用いたデータ分析技法に特化した本を出版することに
 - 年末に出版予定です

なぜ、データが必要
なのだろうか？

たとえば; 経営といえは



ひと



もの



かね



情報

経営を分析する (オールドスクールな方法)



ひと



もの



かね

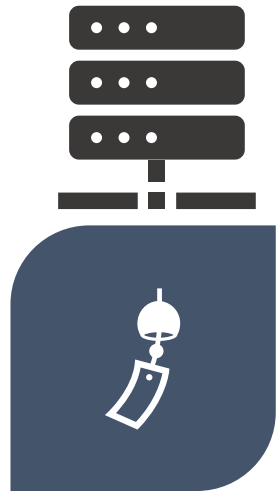


情報



話を聞きに行くか、文献を調査することでなぜその企業/製品がうまくいったのか、要因を経営理論に照らし合わせて解析する(定性的調査)

経営を分析する (最近みんなが好きな手法)



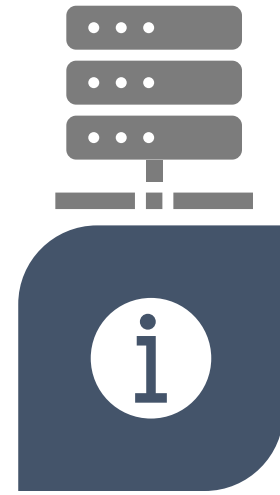
ひと



もの



かね



情報

データベースを入手するか、サーベイデータを作成し、それらのデータセットを接合し合うことで、なぜその企業/製品がうまくいったのか、要因を経営理論に照らし合わせて解析する(定量的調査)

Q. データを使えばビジネスはわかるか？ もしくは、政策評価はできるか？

- ちょっと昔に言われたこと
 - 「データはあくまで数のみをみているだけ。経営者の心の機微や組織体制の細やかな変容や経営者の意思決定の変化を、細かくデータで追うことは出来ない。」
- ここ数年、実際の世の中で起きたこと
 - テキスト解析手法の進化
 - 因果推論の推定手法の開発と普及
 - 機械学習のあつという間の普及
 - データベースの普及と重要性が認識されるようになった
 - データ解析に係る導入コストの低下
 - as 統計ソフトを買わなくても良くなった
- A. 2020年2月段階の(個人的な)答え
 - データがあれば、かなり色々なことがわかる。データ分析の技法は取得したほうが良い。
 - データで説明しきれない特殊性や特筆性については、まだ定性分析に出来ることはある

では、データを使った分析に必要なもの



データ、もしくはデータベース

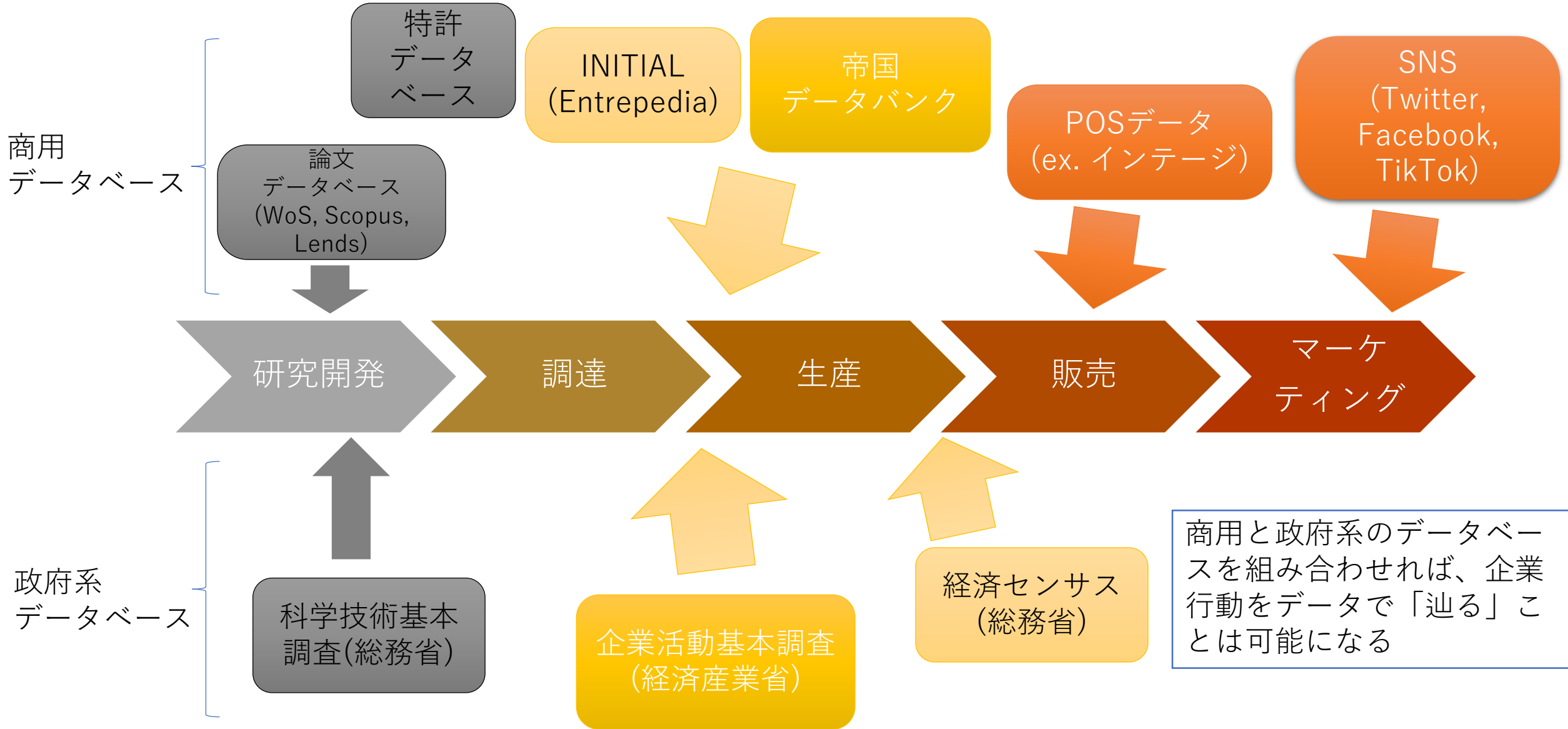


分析手法



分析を行うためのツール

企業活動に関連するデータベース



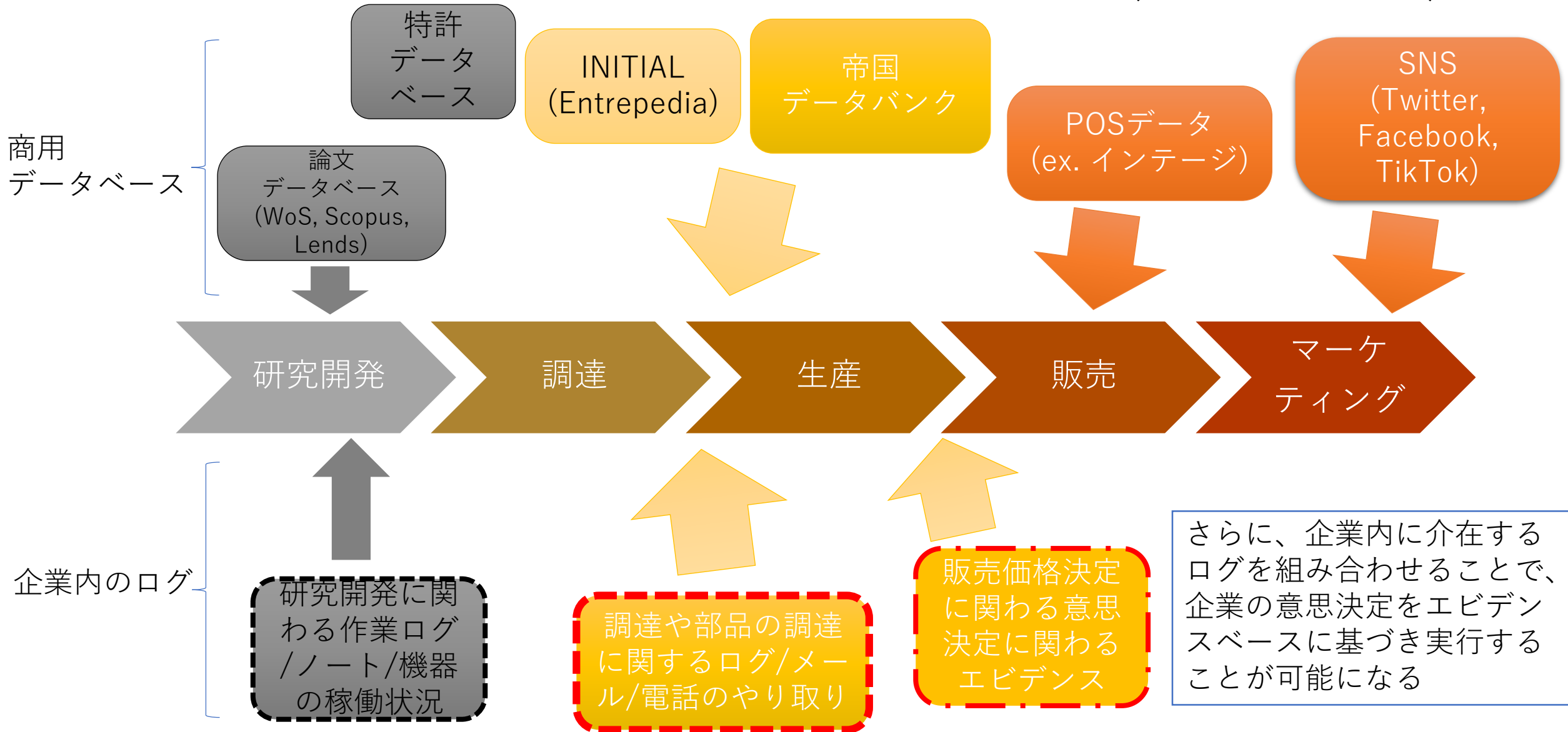
帝国データバンクデータベース (Commercial)

- 一橋CAREE センターで利用可能なデータセット

企業信用調査報告書に関するデータ
取引データ, 出資データ, 銀行取引データ
企業財務データベースに関するデータ
決算書データ
企業概要ファイルに関するデータ
取引データ, 企業情報データ, 銀行取引データ
その他企業に関するデータ
合併データ, 倒産データ

- パネルデータ化が可能.
- アーカイブデータも利用可能(政策フォーラム午前中の岡室先生の発表)
- ご関心のある先生方は、是非ともお問い合わせください

企業活動に関連するデータベース(extended)



データを「データベース化」する

- 不定形データをデータ化する
 - JSON やXML をCSV にして、人間が読めるようにする
 - もしくはその逆
- データをデータベース化する
 - Machine Readable
 - SQL, NoSQL etc...
- データを管理し保持する（データマネジメント）
 - 大昔; 紙
 - 昔; Excel
 - 一昔前; スタンドアロンのサーバ
 - ちょっと昔; パブリッククラウド (AWS, Microsoft Azure, GCP etc...)
 - 現代; サーバーレス

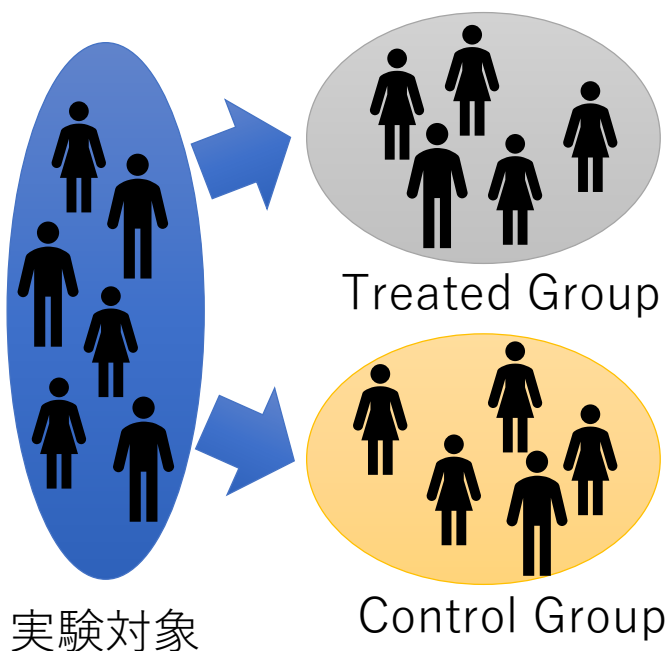
分析手法

- Step1; グラフにする
 - 「まずは散布図を書いてみる」
- Step2; 統計的処理を行う (平均値, 中央値, 最大値, 最小値 etc..)
 - このあたりから、内生性や見せかけの回帰との戦いがはじまる
- Step3; 回帰分析など, 入力と出力間の関係性を見る
 - 重回帰分析
 - スパース推定
 - 決定木分析, SVM (教師あり/教師なしの機械学習)
- Step4; 因果効果を識別する
 - 因果関係と相関関係を識別する(次のページ)

因果識別の手法(1)

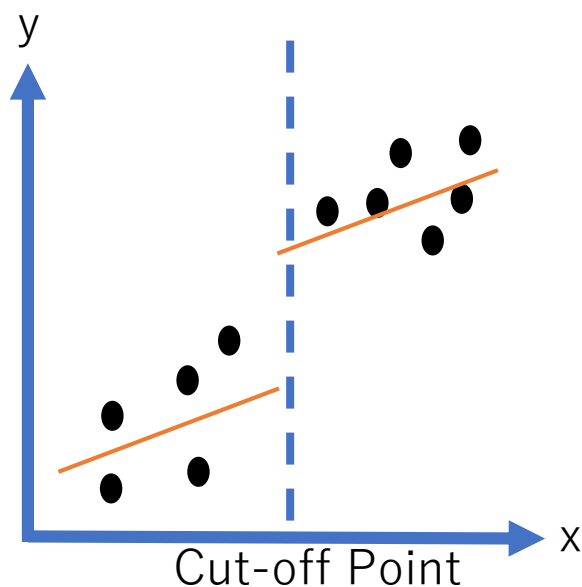
- ランダム化比較試験 (RCT)

- 対象者と非対象者を無作為抽出して比較実験



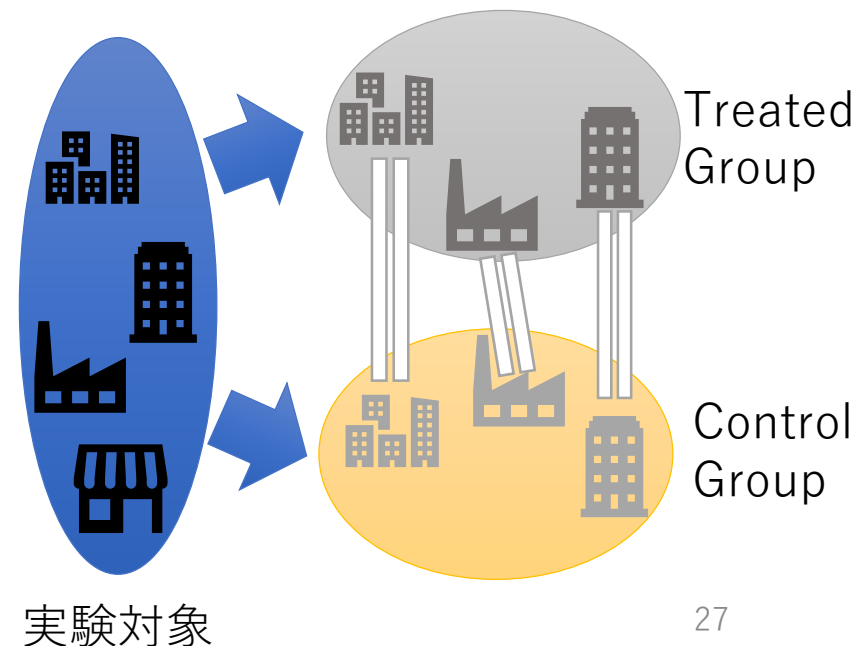
- 回帰不連続(RD) デザイン

- 回帰直線シフト



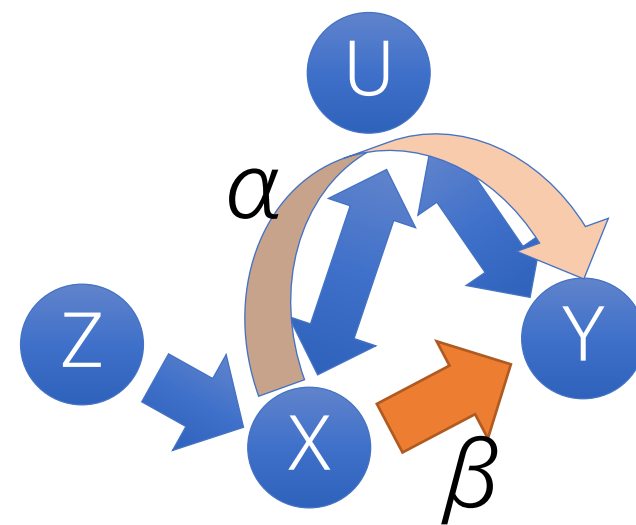
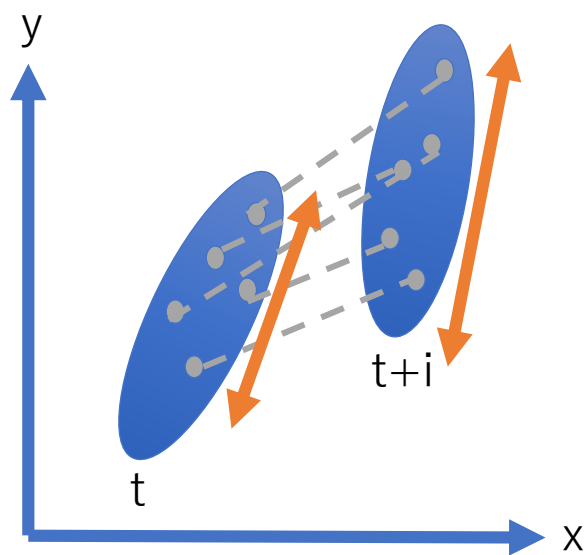
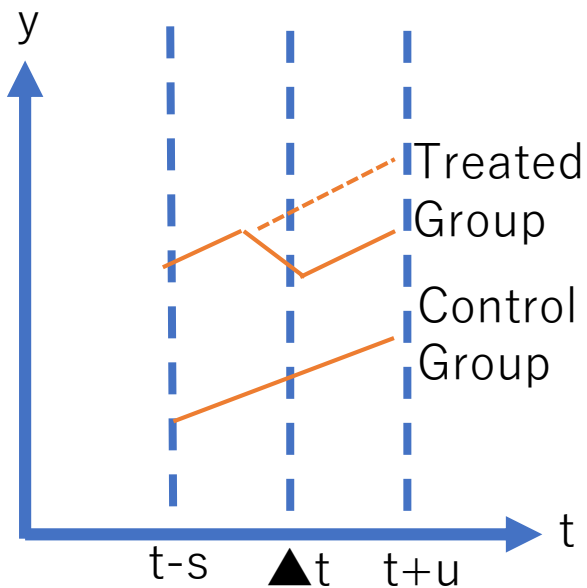
- 傾向スコアマッチング (PSM) :

- 各政策対象企業と同じ「政策対象傾向」を持つ企業を対照群から抽出



因果識別の手法(2)

- 差の差の分析 (DID)
 - 処置群と対照群の比較と政策前後比較の二重差
- パネル固定効果分析
 - パネルデータを用いて、各企業の固有の属性の影響を除去
- 操作変数法 (IV) :
 - 内生変数を外生変数 (操作変数) に置き換える2段階推定

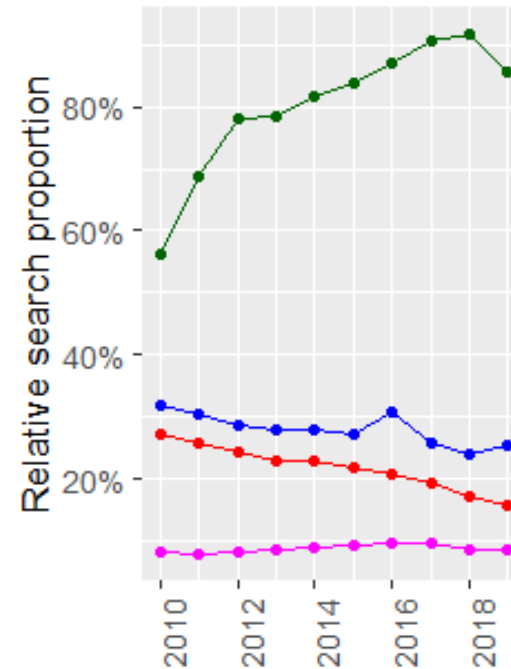


分析ツール

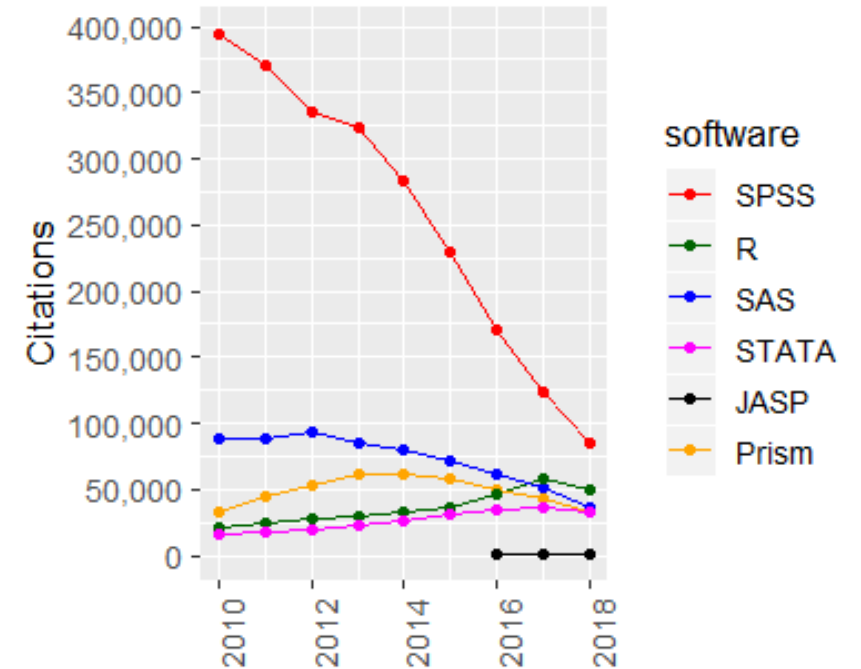
- オープンソース
 - R
 - Python
 - Julia
- 有償ツール
 - Stata
 - SAS
 - SPSS

有償の分析ソフトから、OSS への流れが

Google Trends

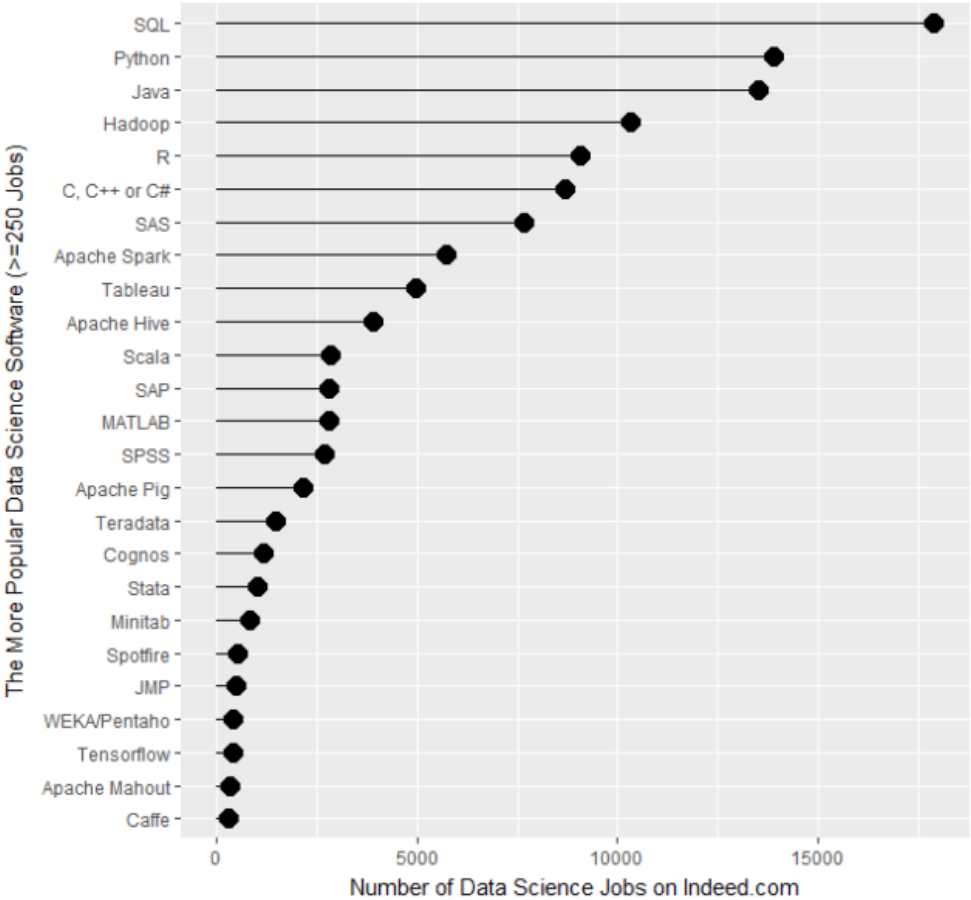


Scholar Citations

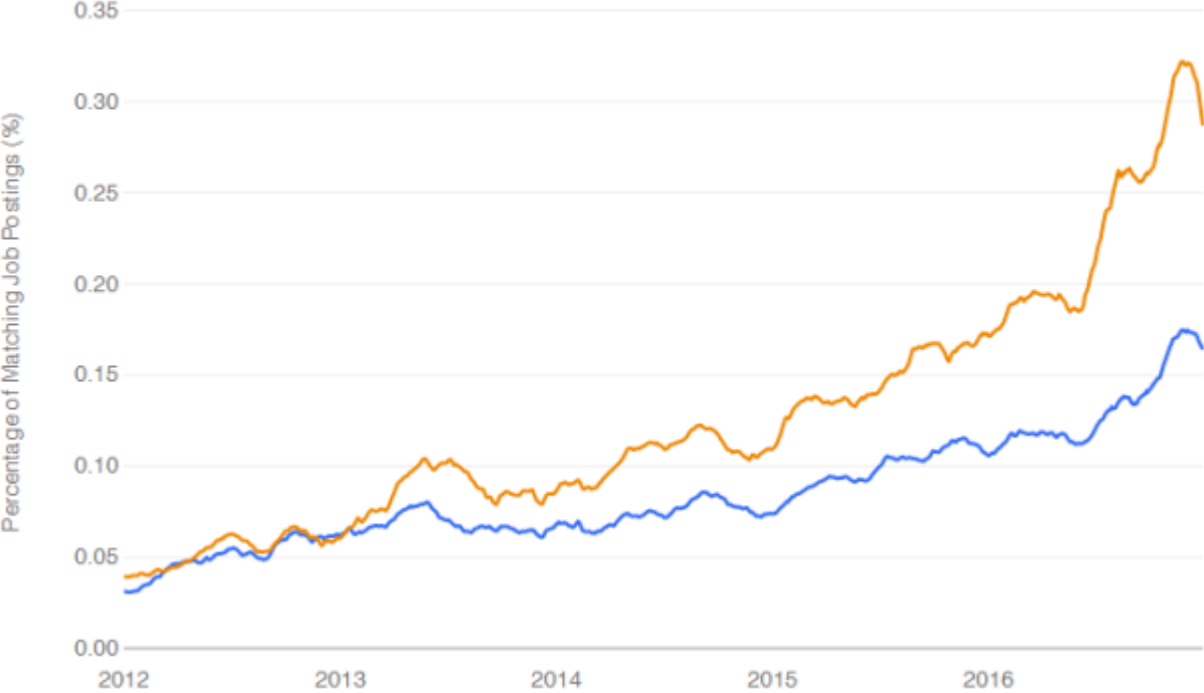


Data Analytics Job Market

The number of data science jobs for the more popular software (those with 250 jobs or more, 2/2017).



Jobs trends for R (blue & lower) and Python (orange & upper).



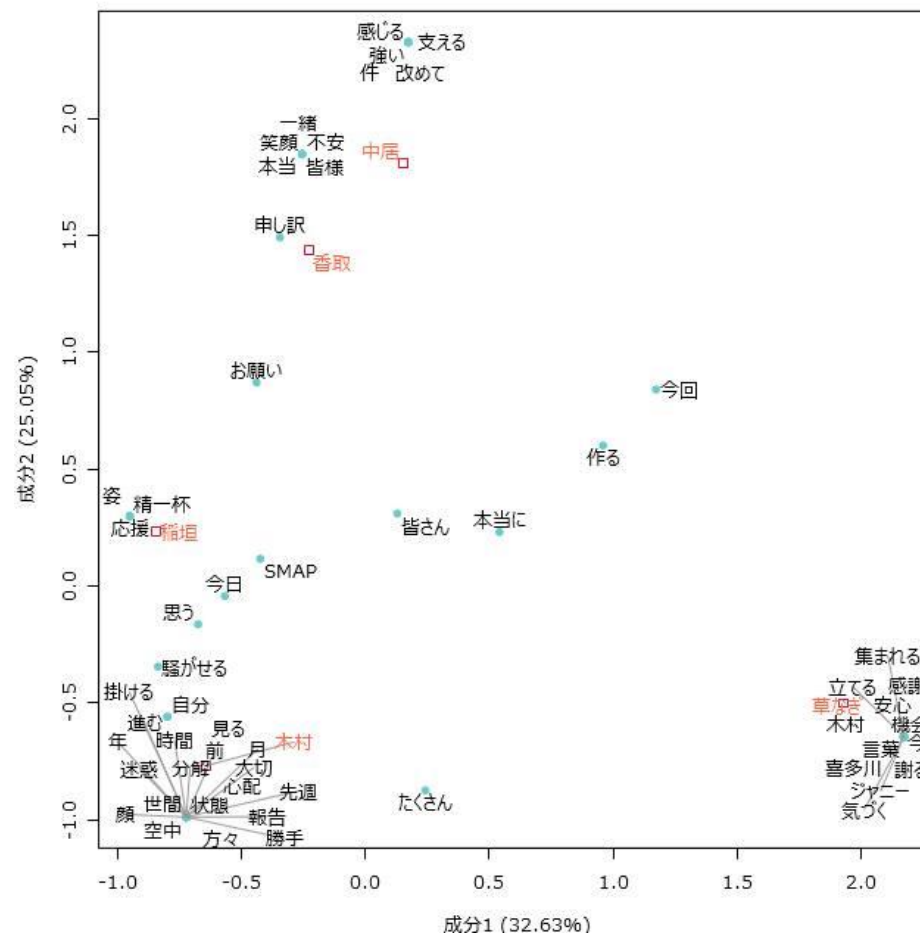
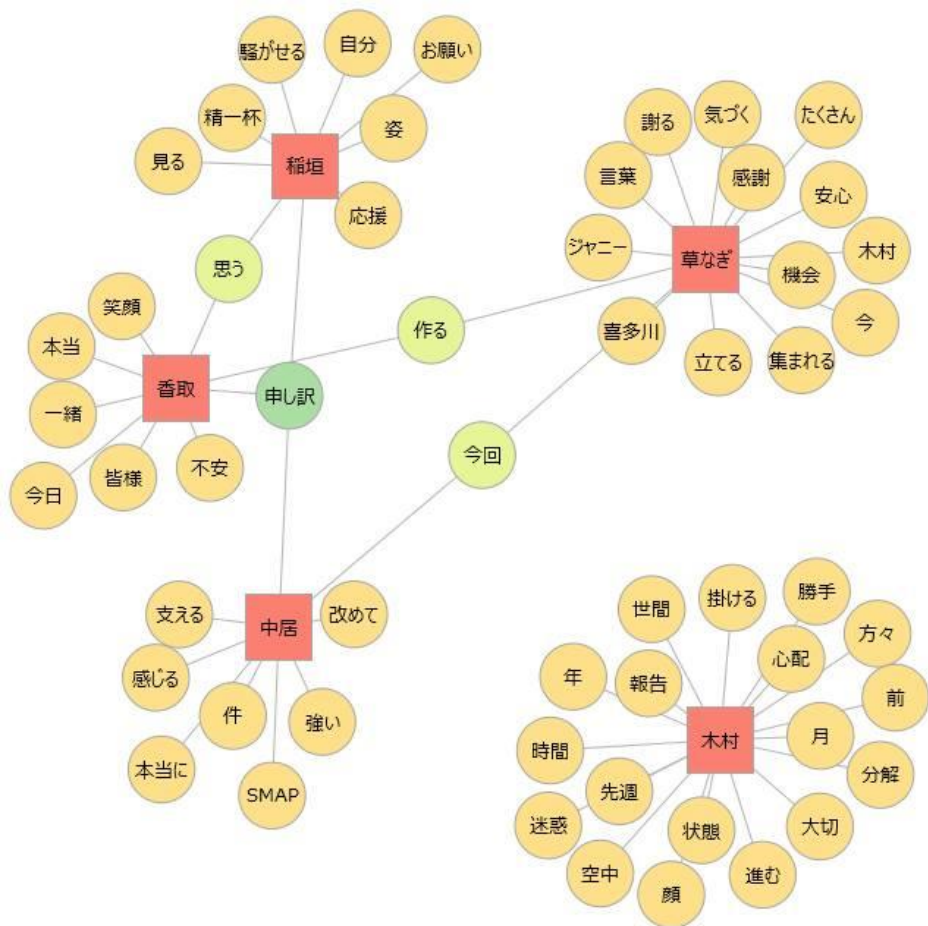
データサイエンスをタピオカミルクティーにしない方法

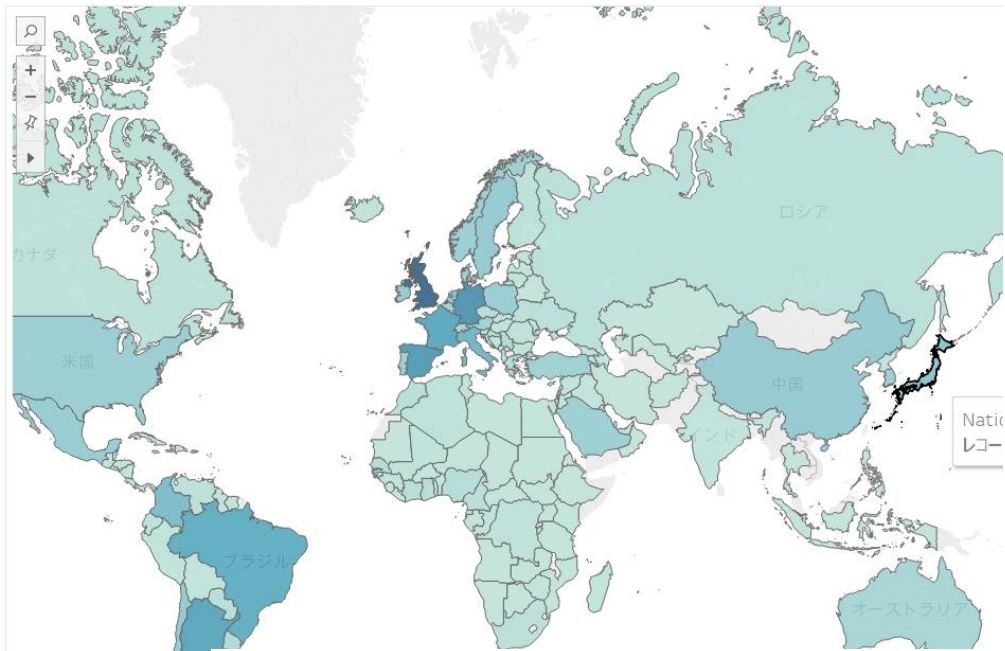
1. データを公開する. Linked Open Data を活用する.
2. データの知的財産権を明確化する.
3. 神エクセルをなくす. (as DX の推進)
4. オープンソースソフトウェアを使う, あるいは教育する.
2020年初頭の最適解はR もしくは Python (3.x系).
5. AI という言葉に踊らされない. 使い倒す.

いくつかの事例

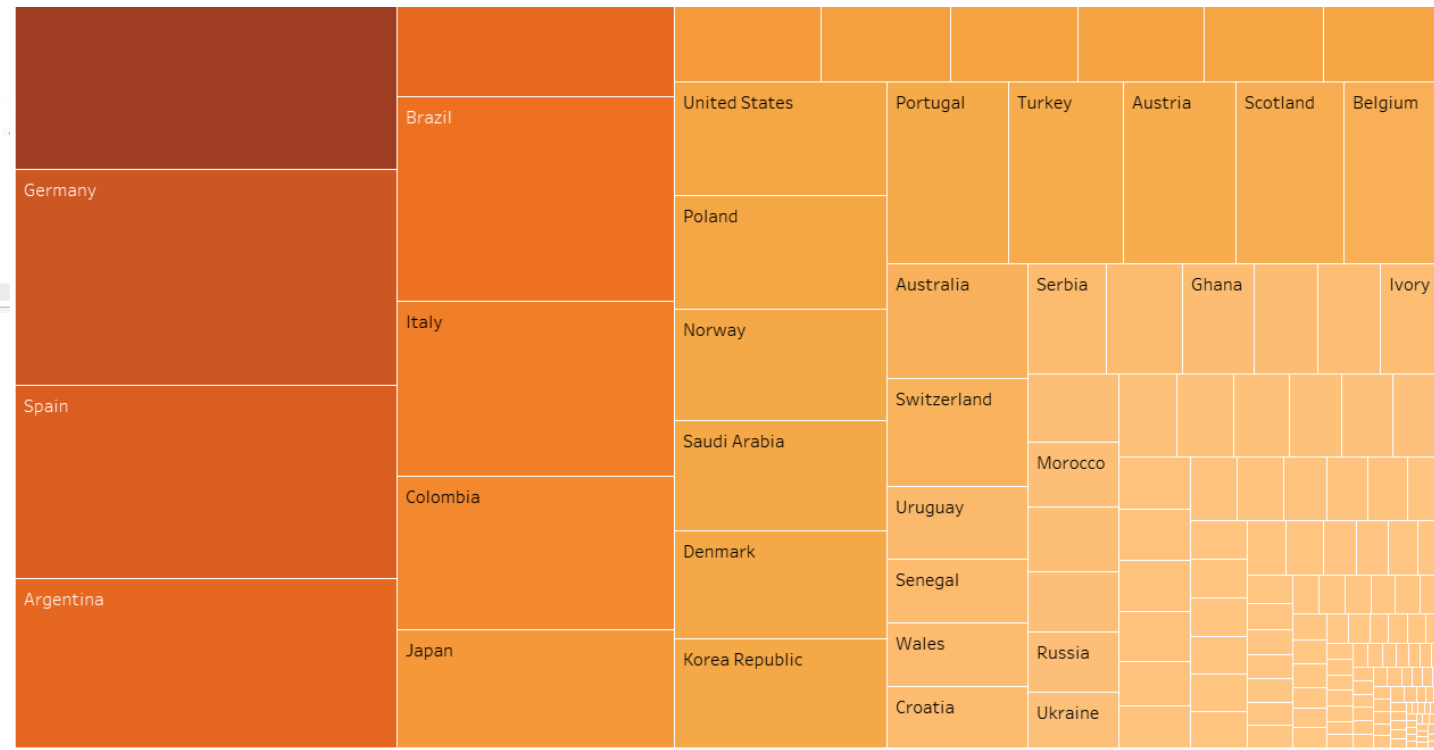
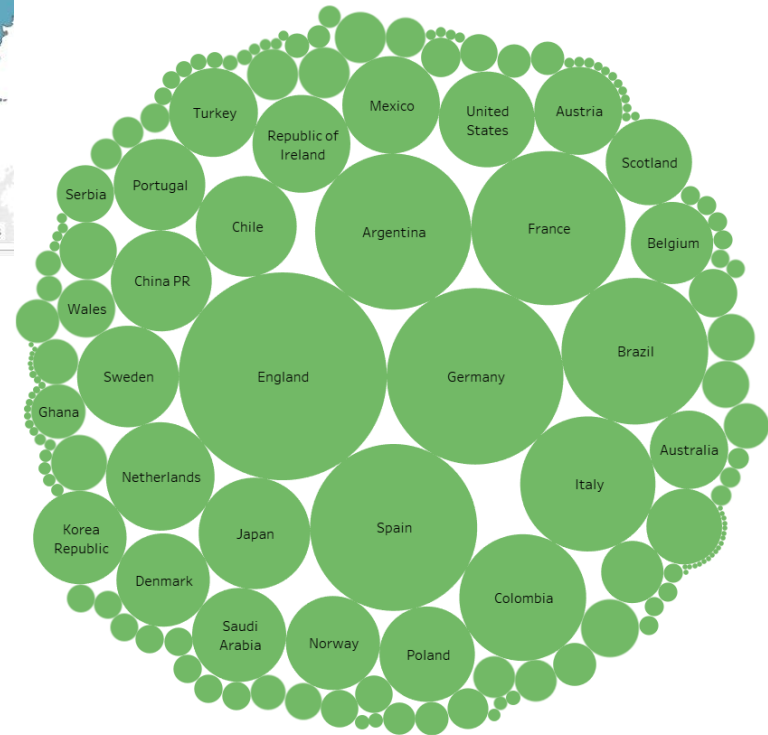
テキスト分析: SMAP会見

- 共起ネットワーク





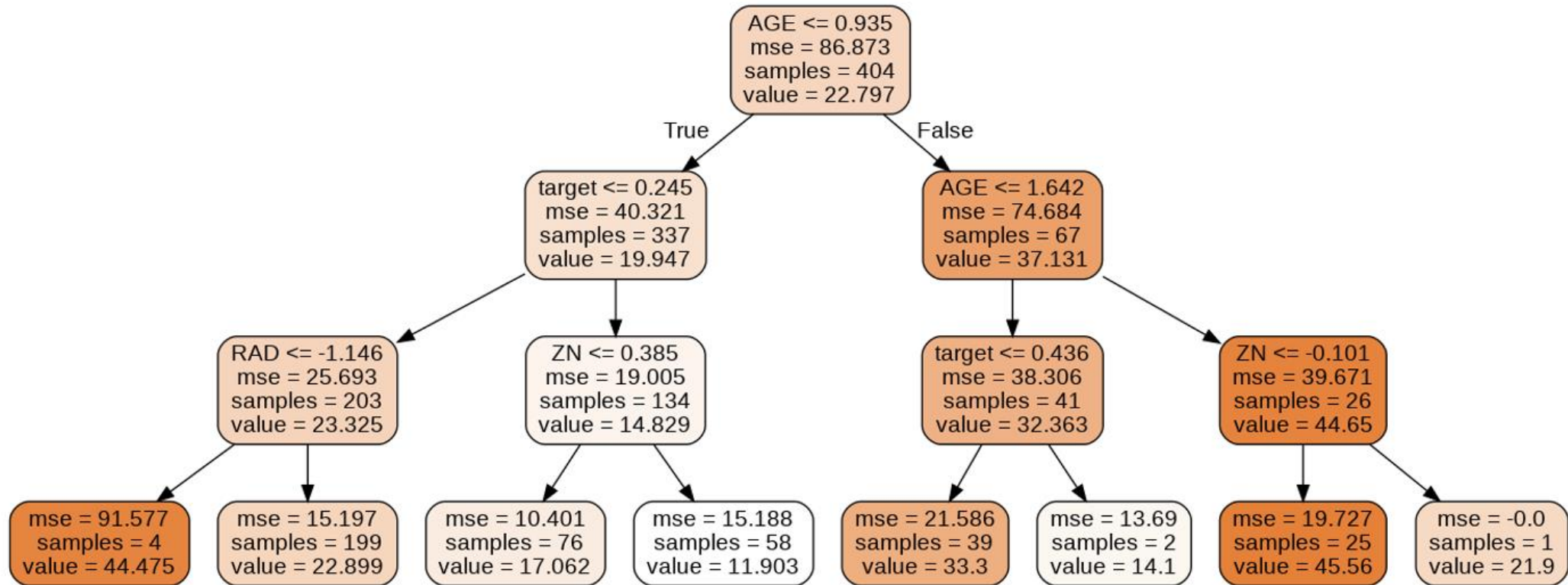
© OpenStreetMap contributors



Word2vec モデルに基づく Similarity Words の抽出 (from Wikipedia の 経営学者リスト)

イノベーション		マーケティング論		技術経営		コーポレート・ガバナンス	
医療	0.92	人的資源	0.92	MOT	0.98	神戸大学大学院経営学研究科	0.93
静岡	0.91	評価	0.92	開発	0.92	第一人者	0.92
地域	0.91	ほか	0.91	システム	0.92	人	0.91
経営情報学部	0.9	勲	0.91	知識	0.91	組織論	0.91
県立大学	0.88	技術	0.9	マネジメント	0.9	および	0.9
センター	0.88	流通	0.9	監事	0.88	消費者	0.89
研究科	0.87	分析	0.9	課題	0.85	にて	0.89
経営情報	0.87	委員会	0.9	Certified	0.85	問題	0.88
研究所	0.84	課題	0.9	客員研究員	0.84	受章	0.88
所長	0.83	人	0.89	技術	0.82	期	0.87
長	0.82	監事	0.89	関係	0.82	現代	0.87
准教授	0.81	および	0.89	プロジェクト	0.82	勲	0.87
国際企業	0.8	多摩大学	0.89	マーケティング論	0.81	アドバイザー	0.87
経済	0.8	地域経済	0.88	領域	0.81	及び	0.87
研究員	0.79	サービス	0.88	ホスピタリティ	0.8	社会学	0.86

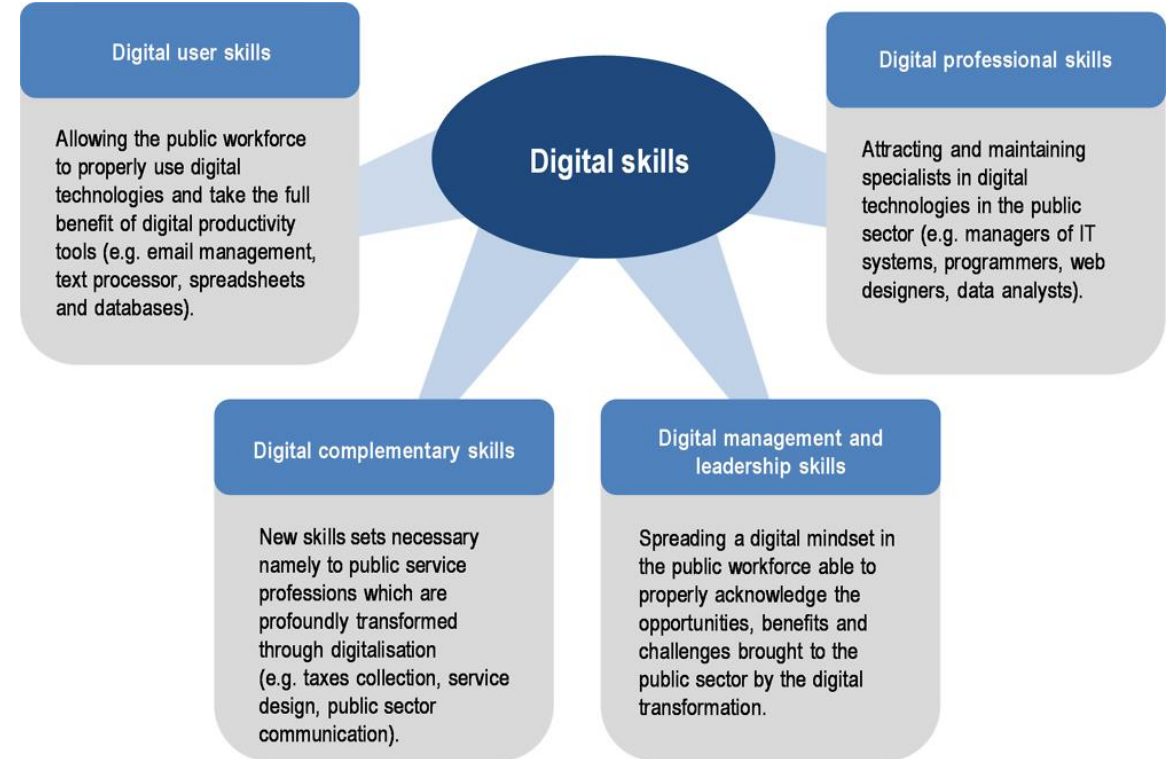
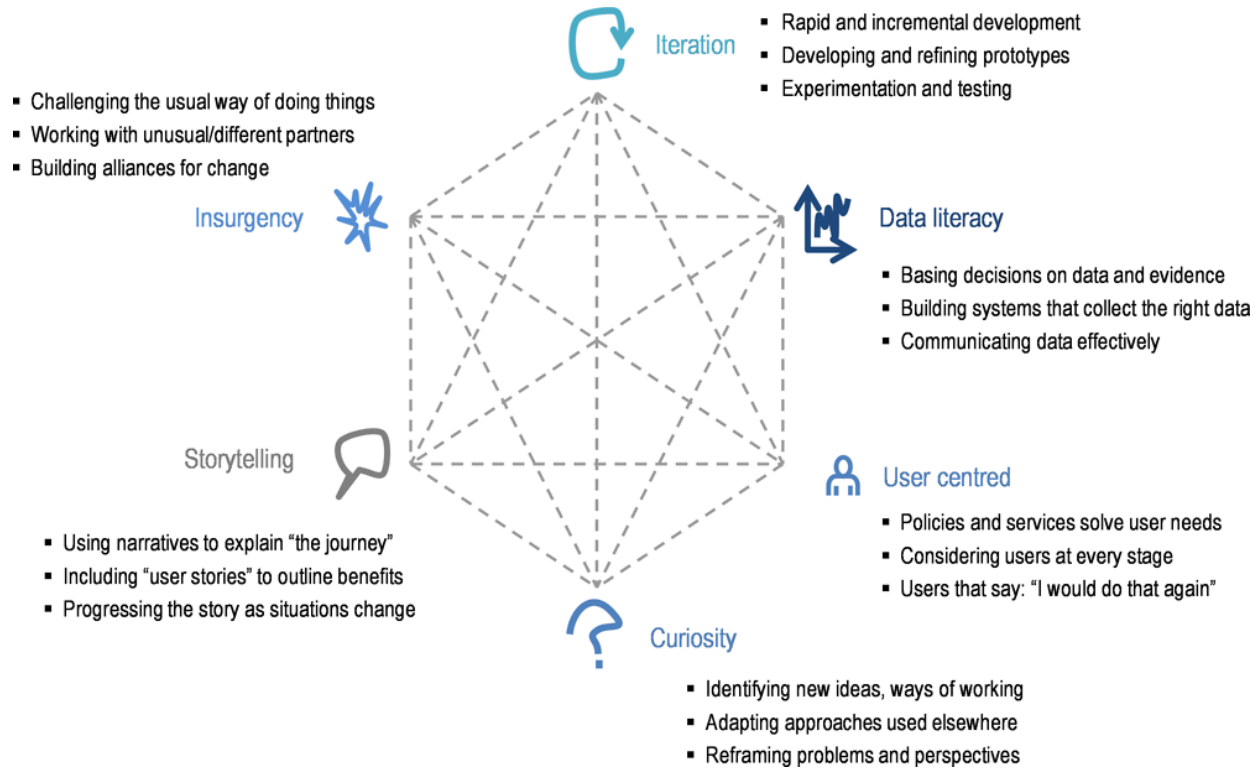
決定木分析



(いくつかの)ギャップ

- 願望投影型政策 (森田 2015) とEBPM
 - データが、推進したい施策に基づき構築される可能性
- ナラティブなストーリーとデータ解析
 - 統計的な素養を政策立案者や企業戦略の決定者すべてが有しているわけではない。データに基づく解析結果を広く共有するためには、ストーリーに落とし込むことが必要
- データのオープン化と、ITセキュリティ対策
 - RESAS や Notebook や Linked Open Data を利用しようとしても、自治体や中央政府、企業のITセキュリティ上制限が掛けられており(ex. 自治体のネットワーク分離), 業務用PCなどでこれらのツールを利用することが出来ない。

OECD が定義するデジタルスキル



Conclusion

- データサイエンスと呼ばれているものは、プログラミングスキルと数学とエンジニアリングスキルと因果関係を特定するためのスキルが混ざりあったもの
- 政府のみなさんへ
 - とりあえずデータを、Machine-Readable な形で公開してください
 - PDF でも Excel でもなくて、JSON やRDF 形式だとかなりベター
- 企業のみなさんへ
 - 組織内のデータと、外部データを組み合わせることで新たな知見を得られる可能性が
 - 経済学や（実証的な規範に基づく）経営学は思ったよりも「役に立つ」と思います

THANKS

yasushi.hara@r.hit-u.ac.jp