



2015年11月20日

プレスリリース  
報道関係各位

一般社団法人データサイエンティスト協会

**データサイエンティストに必要とされるスキルをまとめた  
スキルチェックリストを初公開  
総合人材サービス会社での活用が決定**

一般社団法人データサイエンティスト協会（所在地：東京都港区、代表理事 草野 隆史、以下データサイエンティスト協会）およびスキル委員会（委員長 安宅 和人、副委員長 佐伯 諭）は、データサイエンティストに必要とされるスキルを「データサイエンティストスキルチェックリスト（以下スキルチェックリスト）」としてまとめ、第1版を公開いたしました。また、昨年12月に公開の「データサイエンティストのミッション、スキルセット、定義、スキルレベル」について2015年版を発表いたしました。

データサイエンティストは、ビッグデータやIoTを活用する上で不可欠となる職種といわれており、国内においてもその不足が課題となっております。データサイエンティスト協会は、この課題に対し、データサイエンティストという新しい職業が健全に成長をするための環境整備を目的に2013年より活動を行っております。

活動の一環として昨年12月に発表した、「データサイエンティストのミッション、スキルセット、定義、スキルレベル」は、データサイエンティストとして目指すべき職種像を明らかにしたものでしたが、今回、新たに作成したスキルチェックリストは、各スキルレベルで求められるスキルの内容を具体的に整理したものととなります。

データサイエンティストが有すべきスキルを明確にしたスキルチェックリストがもたらす効果と対象は、以下のとおりです。今後、国内におけるデータサイエンティストの不足と、その定義が曖昧なことにより発生しているさまざまなミスマッチによる社会的損失を解消し、業界の健全な発展に寄与することを期待しております。

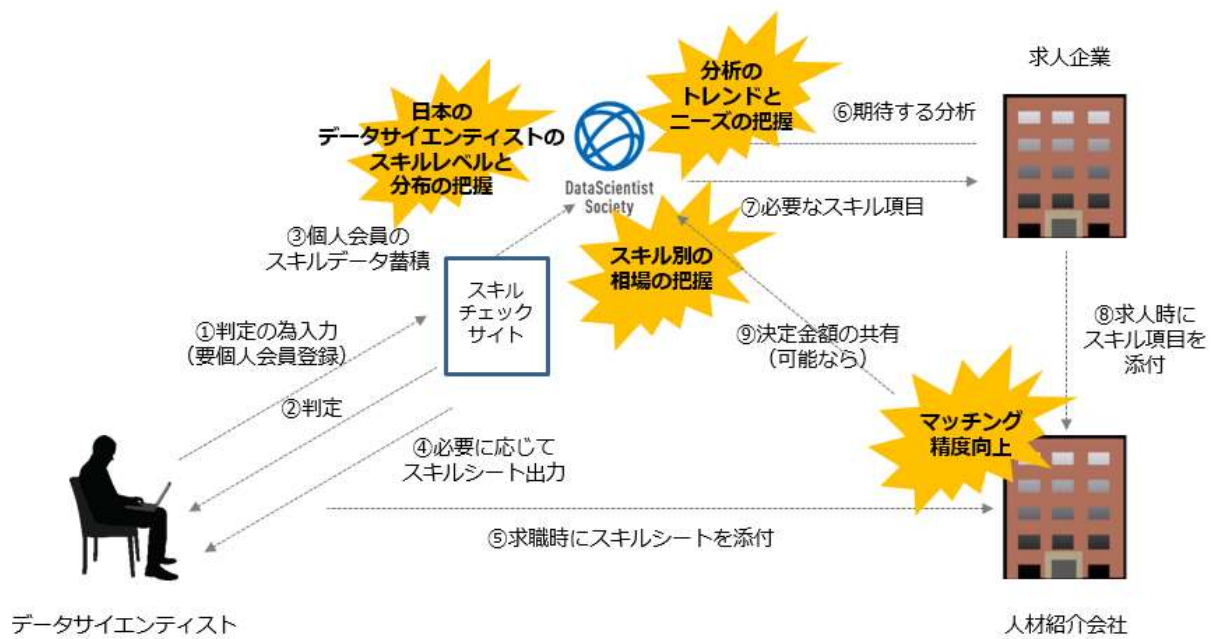
■想定する効果と対象

想定する対象	想定する効果
データサイエンティストを目指す人材	習得すべきスキルの明確化
データサイエンティストを採用する企業	獲得したい人材のスキルの明確化・明文化 同スキルの人材の待遇のバラつき解消
データサイエンティストの求職者	自分のスキルの明示とともに、 期待される仕事のレベルの事前把握
データサイエンティストを紹介する企業	求職者と求人内容のスキルによる マッチング
データサイエンティストを育成する 教育機関・企業	カリキュラムの明確化



データサイエンティスト協会は、スキルチェックリストを利用した自己診断サービス（データサイエンティスト協会のウェブサイト上でチェックすることで自身のスキルレベルを判定したチェックシートが出力される）の提供を来春に予定しており、総合人材サービス会社の株式会社インテリジェンスの人材紹介部門より、データサイエンティスト職の求人求職において同フォーマットを活用したマッチングに賛同の意を表明いただいております。

## ■「データサイエンティストスキルチェックリスト」の展開イメージ



## ■本スキルチェックリストに寄せられた声

株式会社インテリジェンス

キャリア Div.&メディア Div. マーケティング企画統括部

エグゼクティブマネジャー 木下 学 様

株式会社インテリジェンス

キャリアディビジョン マーケティング企画統括部 データアナリティクス部

マネジャー 大江 信明 様

「通常、人材紹介会社では、企業の人事部様に採用要件をヒアリングし、その要件から求人票を作成します。そして転職希望者様とマッチングの上、求人を案内していきます。今回のスキルチェックリストの利用により、企業への採用要件のヒアリングや、転職希望者のスキルのヒアリングにおいて、項目が統一されることにより、マッチングの向上に役立てることができることを期待しております。また、どのように使用したらより有効に機能するのか、引き続き検討したいと考えております。」



## ■スキルチェックリスト公開の背景

### データサイエンティスト協会の設立の背景と目的

データサイエンティスト協会設立時の状況としては、以下の主な課題がありました。

- ・バズワードである「データサイエンティスト」という言葉の定義が欠落
- ・雇い主側の期待に雇われる側のスキルセットが一致しないケースが増加
- ・若い才能達が、自分をどう訓練し、スキルを身に着けていくべきかわからない
- ・企業としてこの新しい職種の適切な処遇や活用の方法が未確立
- ・その他、全般的に情報が不足している

この課題認識を踏まえ、データサイエンティスト協会は、以下の主な目的を果たすために設立されました。

1. この新しいデータプロフェッショナル（「データサイエンティスト」）に必要とされるスキルセットを定義する
2. データサイエンティストの育成と評価のための軸、基準を作る
3. データサイエンティストが相互に接し、情報をやりとりできる環境を提供する
4. 社会に対して、データサイエンティスト及びその仕事を啓蒙するとともに、必要な情報を発信する
5. その他、必要な仕組みづくりを推進する

スキル委員会は、昨年公開した上記の「1. スキルセット」に基づき、今回、「2. データサイエンティストの育成と評価のための軸、基準」の作成検討を進め、得られた成果をスキルチェックリストとしてまとめました。

## ■スキルチェックリストの概要

### データサイエンティストのスキルレベル (2015年版)

データサイエンティストのスキルセット（ビジネス力/データサイエンス力/データエンジニアリング力）別に4段階のスキルレベルを定義しています。この1年間の技術・業界の進展を受けて、スキルレベル表を更新しました。（詳細は、図1を参照）

1. 業界を代表するレベル : Senior Data Scientist (★★★★)
2. 棟梁レベル : Full Data Scientist (★★★)
3. 独り立ちレベル : Associate Data Scientist (★★)
4. 見習いレベル : Assistant Data Scientist (★)

※★の数はスキルレベルを示します。

※独り立ちレベルより上位レベルは、必ずしも一人が全てのスキルを有する必要はないと考えています。全体を俯瞰してコーディネートする人材は必要ですが、個別のスキルに秀でた人とのチームを組成し、実務にあたることも現実的には多いと思っています。



### スキルチェックリストの見方と活用方法

データサイエンティストのスキルセット（ビジネス力／データサイエンス力／データエンジニアリング力）と4段階のスキルレベルにあわせ、各象限におけるデータサイエンティストに必要とされるスキルセットについて、チェックリストを整理しています。

- ・3つの領域別に判定します。
- ・”必須”に○がついているものは、「判定基準を満たしていても、これがないとそのレベルとは認められないもの」として設定しています。
- ・各レベルの判定基準の目安は、「Full Data Scientist」が50%、「Associate Data Scientist」が60%、「Assistant Data Scientist」が70%を想定しています。

スキルチェックリストは、人材育成や組織マネジメントなどで活用可能です。

- ・データサイエンティストのスキルレベルの現状把握や、現状から見た強み・弱みの把握
- ・今後のキャリアアップに向けて、どのような点を強めていけばよいかの把握など
- ・データサイエンス組織のマネジメントにおける上記育成観点や、採用など

### ■データサイエンティストのスキル定義の概要

（2014年12月10日発表 プレスリリース「データサイエンティスト協会、データサイエンティストのミッション、スキルセット、定義、スキルレベルを発表 (<http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>) 」より

データサイエンティストに求められるスキルセット：「ビジネス (business problem solving) 力」、「データサイエンス (data science) 力」、「データエンジニアリング (data engineering) 力」

「データの持つ力を解き放つ」というデータサイエンティストのミッションを踏まえ、データサイエンティストに求められるスキルセットは3つあると考えています。

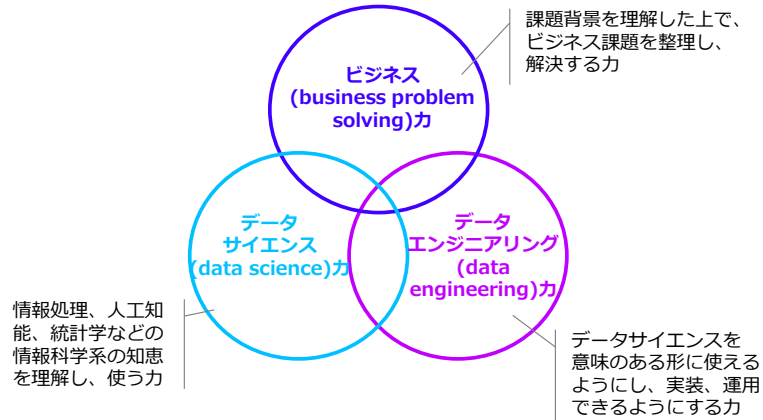
- ・ビジネス (business problem solving) 力：課題背景を理解した上で、ビジネス課題を整理し、解決する力
- ・データサイエンス (data science) 力：情報処理、人工知能、統計学などの情報科学系の知恵を理解し、使う力
- ・データエンジニアリング (data engineering) 力：データサイエンスを意味のある形に使えるようにし、実装、運用できるようにする力



図1：データサイエンティストに求められるスキルセット



## データサイエンティストに求められるスキルセット



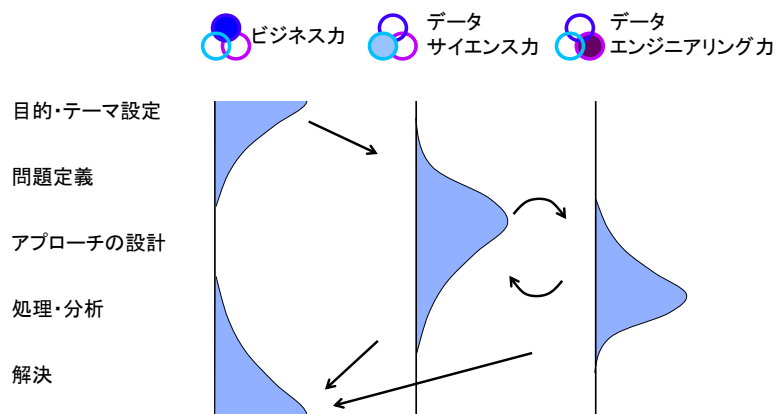
Source: The Japan Data Scientist Society discussions

※この3つのスキルはどの一つが欠けてもいけません。また、この3つのスキルは課題解決のフェーズによって中心となるスキルが変化します。(図2参照)

図2：課題解決の各フェーズで要求されるスキルセットのイメージ



## 課題解決の各フェーズで要求されるスキルセットのイメージ



Source: The Japan Data Scientist Society discussions



### データサイエンティストの定義

必要なスキルセットの検討の結果、これからの時代に求められるデータサイエンティストを次のように定義しました。

「データサイエンティストとはデータサイエンス力、データエンジニアリング力をベースにデータから価値を創出し、ビジネス課題に答えを出すプロフェッショナル」

※ここで「ビジネス」とは社会に役に立つ意味のある活動全般を指します。

※プロフェッショナルとは、体系的にトレーニングされた専門性を持つスキルを持ち、それをベースに顧客（お客様、クライアント）にコミットした価値を提供し、その結果に対して認識された価値の対価として報酬を得る人を示します。



●スキル委員会 委員一覧 (2015年11月現在、敬称略)

	企業名	部署・役職	氏名
委員長	ヤフー株式会社	CSO (チーフストラテジーオフィサー)	安宅 和人
副委員長	株式会社電通	統合データ・ソリューションセンター データ・マネジメント部部长	佐伯 諭
委員	株式会社 ALBERT	取締役会長	山川 義介
	SAS Institute Japan 株式会社	カスタマーサービス本部エデュケーション グループ エデュケーション マネージャ	村上 元親
	株式会社インテリジェンス	キャリアディビジョンマーケティング 企画統括部 データアナリティクス部 マネージャー	大江 信明
	株式会社インテリジェンス	キャリアディビジョンマーケティング 企画統括部 データアナリティクス部	山之下 拓仁
	株式会社タクミインフォメーションテクノロジー	システム開発推進部	原茂 恵美子
	株式会社チェンジ	取締役	高橋 範光
	株式会社大和総研	スキルアップ推進部 次長	松島 寛
	データ・フォアビジョン株式会社	データ・サイエンス本部 本部長補佐	福本 信吾
	デジタル・アドバタイジング・コンソーシアム株式会社	プロダクト開発本部 ビッグデータ解析部 部長	薩摩 定壮
	株式会社電通	統合データ・ソリューションセンター データ・マネジメント部	猪飼 嘉男
	日本アイ・ビー・エム株式会社	テクニカルリーダーシップ 成長イニシアチブ推進 ビジネス・アナリティクス コンサルタント	守谷 昌久
	日本サード・パーティ株式会社	教育コンサルティング部 テクニカルマネージャ	関口 大五郎
	株式会社日立インフォメーションアカデミー	サービス企画部 主任技師	田中 貴博
	株式会社日立インフォメーションアカデミー	ビジネス研修部 技師	大黒 健一
	三菱電機インフォメーションシステムズ株式会社	産業・サービス事業本部 ITコンサルティング部	尾崎 隆
	ヤフー株式会社	データ&サイエンスソリューション統 括本部 データ&サイエンスディレクター	小間 基裕
	ヤフー株式会社	データ&サイエンスソリューション統 括本部 データインフラ本部 本部長	佐々木 潔



- 一般社団法人データサイエンティスト協会について <http://www.datascientist.or.jp/>  
データサイエンティスト協会は、新しい職種であるデータサイエンティストに必要なスキル・知識を定義し、育成のカリキュラム作成、評価制度の構築など、高度 IT 人材の育成と業界の健全な発展への貢献、啓蒙活動を行っています。また、所属を超えてデータ分析に関わる人材が開かれた環境で交流や議論をし、自由に情報共有や意見発信ができる場を提供しています。2015年11月現在、50社6団体の法人会員と約3,050名の一般（個人）会員が参画しています。代表理事：草野 隆史（株式会社ブレインパッド 代表取締役会長）、所在地：東京都港区、設立：2013年5月。

以上

<本件に関するお問い合わせ先>

●本件に関するお問い合わせ

一般社団法人データサイエンティスト協会  
事務局

TEL: 03-6721-9001 e-mail: [info@datascientist.or.jp](mailto:info@datascientist.or.jp)

●報道関係の方からのお問い合わせ

一般社団法人データサイエンティスト協会  
事務局 広報担当

TEL: 03-6721-9001 e-mail: [pr@datascientist.or.jp](mailto:pr@datascientist.or.jp)

\*本プレスリリースに記載されている会社名・商品名は、それぞれの権利者の商標または登録商標です。

\*本プレスリリースに掲載されている情報は、発表日現在の情報です。



## データサイエンティストのスキルレベル

	ビジネス (business problem solving) カ	データサイエンス (data science) カ	データエンジニアリング (data engineering) カ
①Senior Data Scientist 業界を代表するレベル	<ul style="list-style-type: none"> <li>業界を代表するデータプロフェッショナルとして、組織全体や市場全体レベルでのインパクトを産み出すことができる</li> <li>-対象とする事業全体、産業領域における課題の切り分け、論点の明確化・構造化</li> <li>-新たなデータ分析、解析、利活用領域の開拓</li> <li>-組織・会社・産業を横断したデータコンソーシアムの構築、推進</li> <li>-事業や産業全体に対するデータ分析を核としたバリューチェーン創出など</li> </ul>	<ul style="list-style-type: none"> <li>データサイエンスにおける既存手法の限界を打ち破り、新たに課題解決可能な領域を切り拓いている</li> <li>-既存手法では対応困難な課題に対する新規の分析アプローチの開発・実践・横展開</li> <li>-高難度の分析プロジェクトのアプローチ設計、推進、完遂能力など</li> </ul>	<ul style="list-style-type: none"> <li>データサイエンス領域で行おうとしている分析アプローチを、挑戦的な課題であっても安定的に実現できる</li> <li>-複数のデータソースを統合した例外的規模のデータシステム、もしくはデータプロダクトの構築、全体最適化</li> <li>-技術的限界を熟知し、これまでない代案の提示・実行 など</li> </ul>
②Full Data Scientist 棟梁レベル	<ul style="list-style-type: none"> <li>生み出す価値にコミットするプロフェッショナルとして、データサイエンティストとは何かを体現したビジネス判断、課題解決ができる</li> <li>初見の事業領域に向かい合う場合や、スコープが複数の事業にまたがる場合であっても本質的な課題や変数を見出し、構造化し、適切な分析・解析アプローチを設計、実行することができる</li> <li>解決に必要な結果を総合した上で、説得力ある形で共有し、関連する組織、人を動かし、知見の横展開、組織を超えるつなぎ込みができる</li> <li>プロフェッショナルからなる複数のチームによるプロジェクトを推進し、全体としてのアウトプットにコミットできると共に、メンバーを育成、さらには持続的な育成システムを作り出すことができる</li> </ul>	<ul style="list-style-type: none"> <li>予測、グルーピング、機械学習、大量データの可視化、言語処理などの応用的なデータサイエンス関連のスキルを活かし、データ分析プロジェクトの技術的主軸を担うことができる</li> <li>複数もしくは高度な分析プロジェクトを持つチームにおいて、Associate Data Scientist (独り立ちレベル) 以下のメンバーの技能を育成することができる</li> </ul>	<ul style="list-style-type: none"> <li>数十億レコード程度の分析環境の要件定義・設計、データ収集/蓄積/加工/共有プロセスやITセキュリティに関するデータエンジニアリング関連のスキルを活かし、データ分析プロジェクトを中核的に推進することができる</li> <li>複数もしくは高度な分析プロジェクトを持つチームにおいて、Associate Data Scientist (独り立ちレベル) 以下のメンバーの技能を育成することができる</li> </ul>
③Associate Data Scientist 独り立ちレベル	<ul style="list-style-type: none"> <li>大半のケースで自立したプロフェッショナルとしてビジネス判断、課題解決ができる</li> <li>既知の領域、テーマであれば、新規課題であっても解くべき問題の見極めや構造化、その上での分析・解析アプローチの設計、実行、深掘りができる</li> <li>データ、分析結果に対する表面的な意味合いを超えた洞察力を持ち、担当プロジェクトの検討結果を取りまとめ、現場への説明、実装を自律的に行うことができる</li> <li>5名前後のプロフェッショナルによるチームでのプロジェクトを推進しアウトプットにコミットできる</li> <li>-イシュードリブンでスピード感のある判断</li> <li>-プロジェクトマネジメントと個別メンバーの育成</li> <li>-異なるスキル分野の専門家、事業者との協働など</li> </ul>	<ul style="list-style-type: none"> <li>単一プロジェクトにおけるデータ分析をFull Data Scientist (棟梁レベル) に相談しつつ推進できる</li> <li>Assistant Data Scientist (見習いレベル) の日々の活動に適切な指示ができる</li> <li>基礎的な分析活動については、自律的に実施できる</li> <li>-調査目的に合った標本抽出計画の立案</li> <li>-多重 (質問) クロス表を駆使した洞察</li> <li>-パラメトリックな2群の検定の活用 (t検定)</li> <li>-主成分分析や因子分析</li> <li>-適切な学習データとテストデータの作成</li> <li>-重回帰分析を用いた予測モデルの構築</li> <li>-適切な初期値設定を行った非階層クラスター分析</li> <li>-形態素解析などを用いた基本的文書構造解析など</li> </ul>	<ul style="list-style-type: none"> <li>単一プロジェクトにおけるデータ処理・環境構築をFull Data Scientist (棟梁レベル) に相談しつつ推進できる</li> <li>Assistant Data Scientist (見習いレベル) の日々の活動に適切な指示ができる</li> <li>数千レコード程度のデータ処理・環境構築については自律的に実施できる</li> <li>-データの重要性や分析要件に則したシステム要件定義</li> <li>-適切なデータフロー図、論理データモデル作成</li> <li>-Hadoopでの管理対象データ選定</li> <li>-SQLの構文理解と実行</li> <li>-分析プログラムのロジック理解と分析結果検証など</li> <li>データ匿名化方法の理解と加工処理の設計ができる</li> </ul>
④Assistant Data Scientist 見習いレベル	<ul style="list-style-type: none"> <li>ビジネスにおける論理とデータの重要性を理解したデータプロフェッショナルとして行動規範と判断が身についている</li> <li>-データを取り扱う倫理と法令の理解</li> <li>-引き受けたことは逃げずにやり切るコミットメントなど</li> <li>データドリブンな分析的アプローチの基本が身につけており、仮説や既知の問題が与えられた中で、必要なデータを入手し、分析、取りまとめることができる</li> <li>-データや事象のタブリとモレの判断力</li> <li>-分析前の目的、ゴール設定</li> <li>-目的に即したデータ入手と集計、分析</li> <li>-分析の意味合い出しと言語化、取りまとめなど</li> <li>担当する検討領域についての基本的な課題の枠組みを理解、説明できる</li> <li>-ビジネスモデルと主要プレイヤー</li> <li>-基本的なビジネスフレームワークなど</li> </ul>	<ul style="list-style-type: none"> <li>統計数理の基礎知識を有している (代表値、分散、標準偏差、正規分布、条件付き確率、母集団、相関など)</li> <li>データ分析の基礎知識を有している</li> <li>-検定 (帰無仮説、対立仮説・)</li> <li>-分類 (教師あり分類、教師なし分類・)</li> <li>-予測 (回帰係数、標準誤差・)</li> <li>など</li> <li>適切な指示のもとに、以下を実施できる</li> <li>-基本統計量や分布の確認、および前処理 (外れ値・異常値・欠損値の除去・変換や標準化など)</li> <li>-前処理後のデータに対し、抽出、集計、可視化 (ヒストグラム、散布図など)</li> </ul>	<ul style="list-style-type: none"> <li>データやデータベースに関する基礎知識を有している</li> <li>-構造化/非構造化データの判別、論理モデル作成</li> <li>-ER図やテーブル定義書の理解</li> <li>-SDKやAPIの概要理解など</li> <li>数十万件程度のデータ加工技術を有している</li> <li>-ソート、クレンジング、集計、フィルタリングができる</li> <li>-SQLで簡単なSELECT文を記述・実行できる</li> <li>-設計書に基づき、プログラム実装できる</li> <li>適切な指示のもとに、以下を実施できる</li> <li>-同種のデータを統合するシステムの設計</li> <li>-インポート、レコード挿入、エクスポート</li> <li>セキュリティの基礎知識を有している (機密性、可用性、完全性の3要素など)</li> </ul>
DS以前の方	<ul style="list-style-type: none"> <li>ビジネスは勘と経験だけで回すものだと思っている</li> <li>課題を解決する際に、そもそも定量化する意識が無い</li> </ul>	<ul style="list-style-type: none"> <li>基本統計量の意味を正しく理解していない</li> <li>指数を指数で割り算したりする</li> <li>・「平均年収」をそのまま鵜呑みにしたりする</li> <li>グラフ・チャートの使い方が不適切</li> </ul>	<ul style="list-style-type: none"> <li>レポートされてくる数値サマリには目を通すが、特に記憶には残らない</li> <li>アクセス解析システムを使っていない</li> <li>ExcelやAccessは数字しか入れない</li> </ul>

## データサイエンティスト スキルチェックリスト

「データサイエンティスト スキルチェックリスト」は、データサイエンティストに必要とされるスキルをチェックリスト化したものです。チェックリスト内の各項目のスキルレベルは、★の数によって示し、見習いレベル（Assistant Data Scientist）～ 棟梁レベル（Full Data Scientist）までを判定します。

スキルレベル	判定基準
① Senior Data Scientist（業界を代表するレベル） ★★★★★	-
② Full Data Scientist（棟梁レベル） ★★★	★★★の全項目のうち、50%を満たしている。
③ Associate Data Scientist（独り立ちレベル）★★	★★の全項目のうち、60%を満たしている。
④ Assistant Data Scientist（見習いレベル）★	★の全項目のうち、70%を満たしている。

※「必須スキル」に○がついている項目は、判定基準を満たしていても、この項目が達成されていないとそのレベルとは認められない項目として設定しています。

※ 独り立ちレベル以上のレベルは、下位のレベルを満たしていることが前提となります。

### \*引用・改変時の注意事項

- ・チェックリストを引用される場合「一般社団法人データサイエンティスト協会 スキルチェックリストより引用」と引用元を明示願います。その他については著作権法に従っての引用を願います。
- ・チェックリストを改変された場合「一般社団法人データサイエンティスト協会 スキルチェックリストを改変」と改変した旨、明示願います。

スキルカテゴリー一覧							
項目数			項目数				
データサイエンスカ (項目数：180)	1	統計数理基礎	14	データエンジニアリングカ (項目数：119)	1	環境構築	19
	2	予測	16		2	データ収集	12
	3	検定/判断	11		3	データ構造	11
	4	グルーピング	13		4	データ蓄積	16
	5	性質・関係性の把握	11		5	データ加工	13
	6	サンプリング	4		6	データ共有	13
	7	データ加工	8		7	プログラミング	20
	8	Data visualization	36		8	ITセキュリティ	15
	9	機械学習	19	ビジネスカ (項目数：123)	1	行動規範	11
	10	時系列分析	7		2	論理的思考	18
	11	言語処理	10		3	プロセス	20
	12	画像処理	6		4	データの理解・検証	26
	13	音声処理	5		5	データ入手	6
	14	パターン発見	3		6	意味合いの抽出、洞察	10
	16	グラフィカルモデル	3		7	解決	4
	17	統計数理応用	4		8	事業に実装する	8
	18	シミュレーション/データ同化	3		9	活動マネジメント	20
	19	最適化	7		項目数合計		422

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ：チェック項目	必須スキル
1	1	統計数理基礎	★	統計数理基礎	1+4+9+16+25+36をΣを用いて表せる	○
2	2	統計数理基礎	★	統計数理基礎	$\log_a(x)$ の逆関数を説明できる	○
3	3	統計数理基礎	★	統計数理基礎	順列と組合せの式をP、C、m、nを用いて表せる	○
4	4	統計数理基礎	★	統計数理基礎	条件付き確率の意味を説明できる	○
5	5	統計数理基礎	★	統計数理基礎	平均（相加平均）、中央値、最頻値の算出方法の違いを説明できる	○
6	6	統計数理基礎	★	統計数理基礎	母集団データ（3,4,5,7,8）の分散と標準偏差を電卓を用いて計算できる	○
7	7	統計数理基礎	★	統計数理基礎	母（集団）平均が標本平均とは異なることを説明できる	○
8	8	統計数理基礎	★	統計数理基礎	標準正規分布の分散と平均がいくつかわかる	○
9	9	統計数理基礎	★	統計数理基礎	相関関係と因果関係の違いを説明できる	○
10	10	統計数理基礎	★	統計数理基礎	名義尺度、順序尺度、間隔尺度、比例尺度の違いを説明できる	○
11	11	統計数理基礎	★	統計数理基礎	一般的な相関係数（ピアソン）の分母と分子を説明できる	○
12	12	統計数理基礎	★	統計数理基礎	代表的な確率分布の特徴を5つ以上説明できる	○
13	13	統計数理基礎	★	統計数理基礎	変数が量的、質的どちらの場合の関係の強さも算出できる	○
14	1	予測	★	予測	単回帰分析について最小二乗法、回帰係数、標準誤差の説明ができる	
15	2	予測	★★	予測	重回帰分析において偏回帰係数と標準偏回帰係数、重相関係数について説明できる	
16	3	予測	★★	予測	重回帰や判別を実行する際に変数選択手法の特徴を理解し、適用できる	
17	4	予測	★★	予測	ニューラルネットワークの基本的な考え方を理解し、出力される「ダイアグラム」の入力層、隠れ層、出力層の概要を説明できる	
18	5	予測	★★	予測	重回帰分析において多重共線性の対応ができ、適切に変数を評価・除去して予測モデルが構築できる	○
19	6	予測	★★	予測	決定木分析においてCHAID、C5.0などのデータ分割のアルゴリズムの特徴を理解し、適切な方式を選定できる	
20	7	予測	★★	予測	線形回帰分析が量的な予測するのに対して、ロジスティック回帰分析は何を予測する手法か（発生確率予測など）を説明でき、実際に使用できる	
21	8	予測	★★★	予測	予測アルゴリズムに応じ、目的変数と説明変数に対する必要な変数加工処理が設計、実施できる	
22	9	予測	★★★	予測	予測モデル構築において頑健性を維持するための具体的な方法を設計、実施できる	
23	10	予測	★★★	予測	過学習（正則化）の意味を説明ができ、回避する方法の設計・実施ができる	
24	11	予測	★★★	予測	尤度と最尤推定についての説明できる（尤度関数、ネイマンの分解定理、順序統計量）	
25	12	予測	★★★	予測	（モデルに交互作用が含まれていないとき）回帰パラメータとオッズ比の関係について説明できる	
26	13	予測	★★★	予測	ロジスティック回帰分析を行う際に、最小二乗法ではなく最尤法を使う際の利点（回帰誤差が近似的に正規分布しなくても適用できるなど）を説明し、適用することができる	
27	14	予測	★★★	予測	目的（予測・真のモデル推定など）に応じて、適切な損失関数とモデル選択基準（AIC（赤池情報量規準）、BIC（ベイズ情報量規準）、MDL（最小記述長）など）を選択し、モデル評価ができる	
28	15	予測	★★★	予測	データと分析要件から、モデル精度のモニタリング設計・実施と劣化が見込まれるモデルに対するリモデルの設計ができる	
29	16	予測	★★★	予測	分析要件に応じ、量的予測のためのモデリング手法（重回帰、決定木、ニューラルネットワークなど）の選択とパラメータ設定、結果の評価、チューニングを適切に設計・実施・指示できる	
30	1	検定/判断	★	検定/判断	点推定・区間推定の意味を理解している	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ: チェック項目	必須スキル
31	2	検定/判断	★	検定/判断	帰無仮説・対立仮説の意味を理解している	○
32	3	検定/判断	★	検定/判断	第1種の過誤、第2種の過誤、p値、有意水準の意味を理解している	
33	4	検定/判断	★	検定/判断	片側検定、両側検定の意味を理解している	
34	5	検定/判断	★	検定/判断	対応のあるデータ、対応のないデータの意味を理解している	
35	6	検定/判断	★	検定/判断	平均値、分散、平均値の差の検定手法を知っている	
36	7	検定/判断	★★	検定/判断	t検定を理解して、パラメトリックな2群の検定を活用することができる	○
37	8	検定/判断	★★	検定/判断	様々な分散分析の考え方（一元配置、多重比較、二元配置）を理解して、パラメトリックな多群の検定を活用することができる	
38	9	検定/判断	★★	検定/判断	ウィルコクソン検定（マンホイットニーのU検定）を理解して、ノンパラメトリックな2群の検定を活用することができる	
39	10	検定/判断	★★	検定/判断	クラスカル・ウォリス検定を理解して、ノンパラメトリックな多群の検定を活用することができる	
40	11	検定/判断	★★	検定/判断	カイニ乗検定、フィッシャーの直接確率検定を理解して、分割表における群間の関連性の検定を活用することができる	
41	1	グルーピング	★	グルーピング	教師あり分類と教師なし分類の違いを説明できる	○
42	2	グルーピング	★	グルーピング	判別分析とクラスター分析の概要や使い方を説明できる	
43	3	グルーピング	★	グルーピング	階層クラスター分析と非階層クラスター分析の違いを理解している	
44	4	グルーピング	★	グルーピング	クラスター分析のデンドログラムの見方が分かり、適切に解釈できる	
45	5	グルーピング	★★	グルーピング	非階層クラスター分析において、分析目的に合致したクラスター数を決定することができる	
46	6	グルーピング	★★	グルーピング	階層クラスター分析における代表的なクラスター間距離（群平均法、Ward法、最長一致法など）の概念を理解し、分析内容に応じて最適な手法を選択して、分析ができる	
47	7	グルーピング	★★	グルーピング	各種距離関数（ユークリッド距離、マンハッタン距離、cos類似度など）を理解し、分析内容に応じて、適切なものを選択し分析ができる	
48	8	グルーピング	★★	グルーピング	分析内容に応じて、線形か非線形を使い分けた判別分析ができる	
49	9	グルーピング	★★	グルーピング	k-means法は局所最適解であるため初期値問題があることを理解し、適切な初期値を設定して分析を行える	○
50	10	グルーピング	★★★	グルーピング	距離の公理を知っており、距離の公理を満たさない場合（[1-cos類似度]など）のクラスター分析を適切に行える	
51	11	グルーピング	★★★	グルーピング	k-meansの派生手法（x-means、k-means++、ファジィk-meansなど）を理解し、分析内容に応じた適切な手法を選択して、分析ができる	
52	12	グルーピング	★★★	グルーピング	k-meansとカーネルk-means（非線形クラスターリング）、スペクトラルクラスターリングの違いを理解し、分析内容に応じて、適切な手法を選択して、分析ができる	
53	13	グルーピング	★★★	グルーピング	自己組織化マップ（SOM）、Affinity Propagation、混合分布モデル、ディリクレ過程混合モデルを理解し、分析内容に応じて適切な手法を選択して、分析ができる	
54	1	性質・関係性の把握	★	性質・関係性の把握	適切なデータ区間設定でヒストグラムを作成することができる	○
55	2	性質・関係性の把握	★	性質・関係性の把握	何でクロスするかを判断して、表側を設定しクロス集計表を作成することができる	○
56	3	性質・関係性の把握	★	性質・関係性の把握	量的変量の散布図を描き、2変数の関係性を把握することができる	○
57	4	性質・関係性の把握	★★	性質・関係性の把握	多重（質問間）クロス表などを駆使して、データから適切なインサイトを得ることができる	○
58	5	性質・関係性の把握	★★	性質・関係性の把握	主成分分析と因子分析の違いや使い分けを説明できる	○
59	6	性質・関係性の把握	★★	性質・関係性の把握	正準相関分析を説明、活用できる	
60	7	性質・関係性の把握	★★	性質・関係性の把握	正規分布を前提とした多変量解析において、分析の目的と用いるデータの種類から適切な手法を選択できる	○
61	8	性質・関係性の把握	★★	性質・関係性の把握	因子分析における、因子負荷量や因子軸の回転について説明できる	
62	9	性質・関係性の把握	★★	性質・関係性の把握	コレスポンデンス（対応）分析と数量化3類との関係を説明できる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ: チェック項目	必須スキル
63	10	性質・関係性の把握	★★	性質・関係性の把握	コンジョイント分析を用いて効用値と寄与率のグラフを描くことができる	
64	11	性質・関係性の把握	★★	性質・関係性の把握	適切な類似度を設定した上で、多次元尺度構成法を用いてポジショニングマップを描くことができる	
65	1	サンプリング	★	サンプリング	標本誤差とは何かを説明できる	
66	2	サンプリング	★	サンプリング	実験計画法の概要を説明できる	
67	3	サンプリング	★★	サンプリング	分析に最低限必要な標本サイズを設定など、調査目的に合った標本抽出計画を立案できる	○
68	4	サンプリング	★★	サンプリング	属性数と水準数が決まれば適切な直交表を選択し実験計画ができる	
69	1	データ加工	★	データ加工	名義尺度の変数をダミー変数に変換できる	
70	2	データ加工	★	データ加工	標準化とは何かを知っていて、適切に標準化が行える	
71	3	データ加工	★	データ加工	外れ値・異常値・欠損値とは何かそれぞれ知っていて、指示のもと適切に検出と除去・変換などの対応ができる	○
72	4	データ加工	★	データ加工	指示のもと、加工済データに分析上の不具合がないか検証できる	
73	5	データ加工	★★	データ加工	各変数の分布・欠損率などを踏まえて、外れ値・異常値・欠損値の対応を決定できる	○
74	6	データ加工	★★	データ加工	分析要件や各変数の分布などを踏まえて、必要に応じて量的変数のカテゴリサイズを設計・実行できる	
75	7	データ加工	★★	データ加工	加工データに不具合がないか自分でテストを設計し、検証できる	
76	8	データ加工	★★★	データ加工	加工データの統計的な俯瞰によって不具合の早期発見ができるとともに、統計的観点で次ステップの解析に耐えうるデータであるか評価できる	
77	1	Data visualization	★	方向性定義	単に現場の作業支援する場合から、ビッグデータ中の要素間の関連性をダイナミックに表示する場合まで、可視化の目的の広がりについて理解している	
78	2	Data visualization	★★	方向性定義	特異点を明確にする、データ解析部門以外の方にデータの意味を正しく伝える、現場の作業を支援するといった可視化の役割・方向性を判別できる	
79	3	Data visualization	★★★	方向性定義	データ量が膨大で構造が捉えにくい場合や、アウトプットが想像しにくい場合であっても、可視化の役割・方向性を判断できる（ビッグデータ中の要素間の関連性をダイナミックに表現する、細部に入りきらずに問に対して答えを出すなど）	
80	4	Data visualization	★	軸だし	散布図などの軸だしにおいて、縦軸・横軸の候補を適切に洗い出せる	
81	5	Data visualization	★	軸だし	積み上げ棒グラフでの属性の選択など、適切な層化（比較軸）の候補を出せる	
82	6	Data visualization	★★	軸だし	抽出したい意味にふさわしい軸・層化の粒度、順番を考慮して軸のきざみや層化方法を選択できる	
83	7	Data visualization	★★	軸だし	膨大な属性を持つテーブルから目的に有用な属性を選択できる	
84	8	Data visualization	★★★	軸だし	非構造データから分析の軸になりうる候補を抽出し、付加すべき属性候補を適切に出せる	
85	9	Data visualization	★	データ加工	サンプリングやアンサンブル平均によってデータ量を減らして可視化できる	
86	10	Data visualization	★★	データ加工	データの持つ分散量の観点で、高次のデータを主成分分析（PCA）などにより1~3次元のデータに変換できる	
87	11	Data visualization	★★	データ加工	データの持つ特徴を統計量を使って効果的に可視化できる	
88	12	Data visualization	★★★	データ加工	非線形（高次の曲線、渦状の分布など）のデータであっても、高次のデータの次元を、次元圧縮（1~3次元のデータに変換）して、特徴（データの総分散および各データの位置関係）を損なわずに可視化できる	
89	13	Data visualization	★★★	データ加工	ネットワーク構造、グラフ構造などの表現において、ノードとエッジが増えすぎて特徴抽出が困難であっても、データの絞り込みや抽象度を上げることで適切に可視化できる	
90	14	Data visualization	★★★	データ加工	データ量が膨大（ペタバイト以上）なために、処理しきれず描画できない規模のデータに対しても、適度なデータや情報の抽出（間引き）、クラスタリングなどにより可視化しうる状態にデータを加工できる	
91	15	Data visualization	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、特異点の抽出や次元圧縮を通じてデータを圧縮し、リアルタイム表示できる	
92	16	Data visualization	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、データの分割転送、復元を通じて可視化できる	
93	17	Data visualization	★	表現・実装技法	適切な情報濃度（データインク比、データ濃度の概念）を理解し判断できる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ: チェック項目	必須スキル
94	18	Data visualization	★	表現・実装技法	不必要な誇張をしないための軸表現の基礎を理解できている（コラムチャートのY軸の基準点は「0」からを原則とし軸を切らないなど）	
95	19	Data visualization	★	表現・実装技法	強調表現がもたらす効果を理解している（計量データに対しては位置やサイズ表現が色表現よりも効果的など）	
96	20	Data visualization	★	表現・実装技法	1～3次元の比較において目的（比較、構成、分布、変化など）に応じた図表化ができる	
97	21	Data visualization	★	表現・実装技法	端的に図表の変化をアニメーションで可視化できる（人口動態のヒストグラムが経年変化する様子を表現するなど）	
98	22	Data visualization	★	表現・実装技法	データ解析部門以外の方に、データの意味を伝えるサインとしての可視化ができる	
99	23	Data visualization	★★	表現・実装技法	1～3次元の図表を拡張した多変量の比較（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）を適切に可視化できる	
100	24	Data visualization	★★	表現・実装技法	ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる	
101	25	Data visualization	★★	表現・実装技法	GPSデータなどを平面地図上に重ね合わせた可視化ができる	
102	26	Data visualization	★★	表現・実装技法	挙動・軌跡の可視化ができる（店舗内でのユーザの回遊やEye trackingなど）	
103	27	Data visualization	★★	表現・実装技法	適切な情報（意味）を押さええううえで、デザイン性を高めるための要件提示ができる	
104	28	Data visualization	★★★	表現・実装技法	人体、標高を持つ地図、球面などの上にデータを重ね合わせた可視化ができる	
105	29	Data visualization	★★★	表現・実装技法	地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる（多地点での風の動き、飛行物の軌跡など）	
106	30	Data visualization	★	意味抽出	データの性質を理解するために、データを可視化し眺めて考えることの重要性を理解している	
107	31	Data visualization	★	意味抽出	外れ値を見い出すための適切な表現手法を選択できる	
108	32	Data visualization	★	意味抽出	データの可視化における基本的な視点を理解している（特異点、相違性、傾向性、関連性を見出すなど）	
109	33	Data visualization	★★	意味抽出	統計値（代表値の指標、バラツキの指標、有意性の指標、関係式）を正しく読み、回帰式や移動平均線に意味付けできる	
110	34	Data visualization	★★★	意味抽出	分類系の分析において、分布傾向から原因を追究、活用（分類に応じたDM発送による反応率の向上など）、ドリルダウンを計画し主導できる	
111	35	Data visualization	★★★	意味抽出	予測系の分析において、関連性、特異点、変曲点から原因を追究、活用（予測結果に基づく発注管理など）を計画し主導できる	
112	36	Data visualization	★★★	意味抽出	関連系の分析において関連が高い/低い原因、活用（リコメンドなど）、ドリルダウンを計画し主導できる	
113	1	機械学習	★	機械学習	機械学習にあたる解析手法（Random Forestなど）の名称を3つ以上知っており、手法の概要を理解している	
114	2	機械学習	★	機械学習	指示を受けて機械学習のモデルを使用したことがあり、どのような問題を解決することができるか理解している	
115	3	機械学習	★★	機械学習	「教師あり学習」「教師なし学習」「強化学習」の違いを理解しており、データの内容や学習手法に応じて適切な学習データとテストデータを作成できる	○
116	4	機械学習	★★	機械学習	決定木とRandomForestの違いを説明できる	
117	5	機械学習	★★	機械学習	ROCカーブを用いてモデルの精度を評価できる	
118	6	機械学習	★★	機械学習	混同行列（正誤分布のクロス表）を用いてモデルの精度を評価できる	
119	7	機械学習	★★	機械学習	行列分解（非負値行列因子分解、特異値分解）をツールを使って実行でき、その結果を正しく解釈できる	
120	8	機械学習	★★	機械学習	サポートベクターマシンによる分析を、統計解析ツールを使って実行でき、その結果を解釈できる	
121	9	機械学習	★★	機械学習	アンサンブル学習（Random Forest、勾配ブースティングマシン、バギング）による分析を、統計解析ツールを使って実行でき、その結果を正しく解釈できる	
122	10	機械学習	★★	機械学習	ニューラルネットワークによる分析を、統計解析ツールを使って実行でき、その結果を正しく解釈できる	
123	11	機械学習	★★★	機械学習	行列分解（非負値行列因子分解、特異値分解）を、目的に応じてパラメータを最適化して分析できる	
124	12	機械学習	★★★	機械学習	課題やデータ型に応じて、サポートベクターマシンの適切なモデルを選定し、目的に応じてアルゴリズムの調整や設計ができる	
125	13	機械学習	★★★	機械学習	課題やデータ型に応じて、アンサンブル学習（Random Forest、勾配ブースティングマシン、バギング）の適切なモデルを選定し、目的に応じてアルゴリズムの調整や設計ができる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ: チェック項目	必須スキル
126	14	機械学習	★★★	機械学習	Deep Learning (深層学習) の適切なモデルを選定し、目的に応じてアルゴリズムの調整や設計ができる	
127	15	機械学習	★★★	機械学習	モデル選択のための汎化能力評価 (汎化誤差の推定) を、ホールドアウト法、交差確認法、一つ抜き法、ブートストラップ法を用いて行える	
128	16	機械学習	★★★	機械学習	正則化 (誤差関数に正則化項を追加することで、モデルの複雑度・自由度を抑制する) を利用して、過学習を防ぐことができる	
129	17	機械学習	★★★	機械学習	高次元データの取り扱いについて、次元の呪いを考慮し適切に次元削減できる	
130	18	機械学習	★★★	機械学習	機械学習等の最新の論文を理解し、必要とあれば自分で実装し追試できる	
131	19	機械学習	★★★	機械学習	バイアスとバリエーションの関係を理解し、モデル選定を適切に行える	
132	1	時系列分析	★	時系列分析	時系列データについて説明ができる (時系列グラフ、周期性、移動平均など)	
133	2	時系列分析	★★	時系列分析	時系列データの時点差での相関関係を、系列相関やコレログラムを利用して評価ができる	
134	3	時系列分析	★★	時系列分析	時系列データに対し、ツールを使用して、分析結果の比較を行い、適切なモデルを選択できる (自己回帰モデル、統合モデル、移動平均モデル、ARIMA、SARIMAなど)	
135	4	時系列分析	★★	時系列分析	時系列分析の特徴 (通常は、トレンドおよび季節成分を省いたランダム部分を、定常時系列として分析する手法であること) を理解している	
136	5	時系列分析	★★★	時系列分析	状態空間モデルにおいて、カルマンフィルタを用いて、観測値から欠測値の補間をし、予測モデルを構築できる	
137	6	時系列分析	★★★	時系列分析	非線形・非ガウス型状態空間モデルにおいて、モンテカルロ・フィルタを用いて、複雑な時系列システムの予測モデルを構築できる	
138	7	時系列分析	★★★	時系列分析	ピリオドグラムにおいて、FFT (Fast Fourier Transform) など計算量を抑制する方法により、ピリオドグラムの計算ができる	
139	1	言語処理	★	言語処理	形態素解析や係り受け解析の概念を理解している	
140	2	言語処理	★★	言語処理	形態素解析や係り受け解析のツールを適切に使い、基本的な文書構造解析を行うことができる	○
141	3	言語処理	★★	言語処理	TF-IDFやcos類似度などの基本的なアルゴリズムを使い、単語ベクトルの作成や文書群の類似度計算を行うことができる	
142	4	言語処理	★★	言語処理	Precision、Recall、F値といった評価尺度を理解し、実際の精度評価を行うことができる	
143	5	言語処理	★★★	言語処理	形態素解析・構文解析・固有表現抽出のアルゴリズムを理解し、使いこなせる	
144	6	言語処理	★★★	言語処理	N-gram言語モデルの構築方法と代表的なスムージングアルゴリズムを理解し、使いこなせる	
145	7	言語処理	★★★	言語処理	索引型の全文検索の仕組み (転置インデックス、スコアリング、関連性フィードバック) を理解し、使いこなせる	
146	8	言語処理	★★★	言語処理	Trie、Suffix Arrayなどの代表的な高速文字列検索アルゴリズムを理解し、使いこなせる	
147	9	言語処理	★★★	言語処理	潜在的意味解析 (LSA) の仕組みを理解し、使いこなせる	
148	10	言語処理	★★★	言語処理	データの特性に合わせ、適切な言語処理アルゴリズムを選択し、誤り分析、辞書作成などを行い、成果を最大化することができる	
149	1	画像処理	★	画像処理	画像映像のデジタル表現の仕組みと代表的な画像映像フォーマットの概念を理解している	
150	2	画像処理	★★	画像処理	主要な画像映像フォーマットの種類および特性を理解し、適切に使い分けたり変換することができる	
151	3	画像処理	★★★	画像処理	画像に対して、目的に応じた適切な色変換や補正、フィルタ処理などを行うことができる	
152	4	画像処理	★★★	画像処理	問題に合わせてパターン検出や画像特徴抽出などの既存手法から適切なものを選択し、画像の処理や解析に用いることができる	
153	5	画像処理	★★★	画像処理	オプティカルフローやトラッキング手法などにより映像中の人や物体の移動を分析することができる	
154	6	画像処理	★★★	画像処理	目的に合わせてショット検出、物体検出、行動認識などの手法を用い、映像中の特定のイベントを効率よく検出することができる	
155	1	音声処理	★	音声処理	代表的な音声フォーマットの概要を理解している	
156	2	音声処理	★★	音声処理	代表的な音声フォーマットを理解し、波形データの抽出をすることができる	
157	3	音声処理	★★	音声処理	代表的な音声処理分野の概念を理解し、説明することができる	
158	4	音声処理	★★★	音声処理	代表的な音声信号分析手法を理解し、使いこなすことができる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データサイエンスカ: チェック項目	必須スキル
159	5	音声処理	★★★	音声処理	目的に合わせて、音声認識のパラメータ調整や手法変更、言語モデル・音響モデルの差し替えができ、音声認識の精度評価をすることができる	
160	1	パターン発見	★★	パターン発見	条件Xと事象Yの関係性をリフト値を用いて評価できる	
161	2	パターン発見	★★	パターン発見	アプリアリアルゴリズムのアソシエーション分析の入力パラメータ値を調整して出力件数を調整できる	
162	3	パターン発見	★★★	パターン発見	レコメンドーション業務の要件から適切な協調フィルタリングロジック（ユーザベース、アイテムベースなど）の選定を行うことができる	
163	1	グラフィカルモデル	★★	グラフィカルモデル	パス解析において、変数間の因果関係をパス図を用いて説明できる	
164	2	グラフィカルモデル	★★	グラフィカルモデル	ベイジアンネットワーク分析結果のネットワーク図から目的事象の同時確率を算出できる	
165	3	グラフィカルモデル	★★	グラフィカルモデル	ツールを用いて共分散構造分析（構造方程式モデリング: SEM）を行い、因果関係を説明できる	
166	1	統計数理応用	★	統計数理基礎	ベイズの定理を理解している	
167	2	統計数理応用	★	統計数理応用	ポロノイ図の概念と活用方法を説明できる	
168	3	統計数理応用	★★★	統計数理応用	空間的自己相関の手法を用いて空間的な類似性を数値化できる	
169	4	統計数理応用	★★★	統計数理応用	当該分野に則したベイズ統計に基づくアルゴリズムを理解し、モデルを構築できる	
170	5	統計数理応用	★★★	統計数理応用	ベイズ統計と頻度論による従来の統計との違いを、尤度、事前確率、事後確率などの用語を用いて説明できる	
171	1	シミュレーション/ データ同化	★★	シミュレーション/ データ同化	離散型・連続型シミュレーションについて、説明できる（モンテカルロ、ヒストリカル、Agentベースなど）	
172	2	シミュレーション/ データ同化	★★★	シミュレーション/ データ同化	データ同化（データを用いてシミュレーション内の不確実性を減少させる計算技法）の概念を理解している	
173	3	シミュレーション/ データ同化	★★★	シミュレーション/ データ同化	シミュレーションにおける問題を理解し、対処を考えることができる（初期条件・境界条件・パラメータの不確実性、データ分布の不均一性、実験計画の最適性など）	
174	1	最適化	★★	最適化	線形計画法について、説明することができる	
175	2	最適化	★★	最適化	連続最適化（制約なし）において、アルゴリズムを説明することができる（ニュートン法、最急降下法など）	
176	3	最適化	★★	最適化	制約がある最適化問題において、代表的な組み合わせ解法の概念を説明することができる（一般的な手法（Brute-force search、greedy algorithmなど）、ヒューリスティック（局所探索、山登りなど）、メタ解法（GA、TS、SAなど））	
177	4	最適化	★★★	最適化	代表的な最適化問題に関して、モデリングを行い、ソルバーを使い、最適化できる（ナップザック問題、ネットワークフロー問題、巡回路問題など）	
178	5	最適化	★★★	最適化	ビジネス課題にあわせて、変数、目的関数、制約を定式化し、線形・非線形を問わず、最適化モデリングができる	
179	6	最適化	★★	最適化	凸関数とは何かを説明でき、目的関数が凸関数で実行可能な領域が凸集合である場合の数値計画法を凸計画法ということを理解している	
180	7	最適化	★★	最適化	複数のA/Bテストの統計的結果を踏まえ、デザイン等の最適化を行う手法を回すことができる	



No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データエンジニアリングカ：チェック項目	必須スキル
1	1	環境構築	★	システム運用	サーバー1~10台規模のシステム構築、システム運用を指示書があれば実行できる	
2	2	環境構築	★	システム運用	数十万レコードを持つデータベースのバックアップ・アーカイブ作成など定常運用ができる	
3	3	環境構築	★★	システム運用	数千万レコードのデータを保持するシステムにおいてデータの重要性や分析要件に則したシステム構築、初期データ投入方法、システム運用の要件定義が行える	○
4	4	環境構築	★★	システム運用	顧客管理など分析システムの運用（異常検知、フェイルオーバー、バックアップ、リカバリ処理、開始・停止処理）の手順書作成や要件定義が行える	
5	5	環境構築	★★	システム運用	数千万レコードのデータを保持するシステムのキャパシティ要件（データ容量）と必要処理性能（スループット）を定義できる	
6	6	環境構築	★★★	システム運用	数十億レコードのデータに対してリバランシングなども含めてシステム拡張方法や最適化の要件を整理できる	
7	7	環境構築	★★★	システム運用	扱うデータのデータ規模や機密性、分析要件を理解した上で、オンプレミスで構築するか、クラウド上で構築するかの要件を整理できる	
8	8	環境構築	★	システム企画	データベースから何らかのデータ抽出方法を活用し、小規模なExcelのデータセット作成ができる	○
9	9	環境構築	★	システム企画	オープンデータ活用目的でExcelを使った分析システムの要件定義が行える	
10	10	環境構築	★★	システム企画	HTTPを活用したオープンAPIと分析システムのサーバー環境及びデータベースの連携設計ができる	
11	11	環境構築	★★	システム企画	社内分析者向けのRDBMS、NoSQL、ETL、Visualizationなど単一コンポーネントのユーザー利用機能設計が行える	○
12	12	環境構築	★★	システム企画	ソフトウェア開発プロジェクトの管理方法、設計、テスト方法を理解した上で、データ管理・分析システムを要求定義することができる	
13	13	環境構築	★★★	システム企画	数十億レコードのデータに対してHiveとImpala、Sparkを組み合わせたなど処理可能なデータ量の違うシステムを連携させ、処理性能とのトレードオフを調整できる	
14	14	環境構築	★★★	システム企画	数十億規模の内部データ（CRM、webログ、ユーザー購買データ）、外部データ（購入しているデータ、オープンデータ）を理解し、複数のデータソースを統合する要件を整理できる	
15	15	環境構築	★★★	システム企画	サービス上のそれぞれの機能がどのデータに関連があるか把握し、分析機能追加やシステム変更の要件を整理できる	○
16	16	環境構築	★★★	システム企画	KVS、カラム指向、ドキュメント指向などデータ構造の異なる複数のシステムからデータ取得と分析環境への連携が設計できる	
17	17	環境構築	★★★	システム企画	数十億規模のデータを持つ業務要件やリソース負荷に応じて、データフローや管理機構の統合、またバッチ実行スケジュールの最適化について要件定義が行える	
18	18	環境構築	★★★	アーキテクチャ設計	全体システム化計画及び個別システム化構想・計画を具体化するために、全体最適の観点を持ちながら、対象とするデータ分析システムの開発に必要な要件を整理することができる	
19	19	環境構築	★★★	アーキテクチャ設計	データ活用戦略を正しく理解し、各種業務プロセスについての専門知識とシステムに関する知識を有し、双方を活用して、適切な要求定義が行える	
20	1	データ収集	★	クライアント技術	対象プラットフォームが提供する機能（SDKやAPIなど）の概要を説明できる	○
21	2	データ収集	★★	クライアント技術	対象プラットフォーム（iOS、Android、HEMSなど）におけるデータ取得の仕様（精度など）を理解しており、システム要件を満たせるか判断できる	
22	3	データ収集	★★	クライアント技術	目的に適したログ取得項目を、対象プラットフォーム（iOS、Android、HEMSなど）で取得可能なデータを用いて設計できる	
23	4	データ収集	★★	クライアント技術	対象プラットフォームにおけるバッテリー消費や通信速度などを含めたシステム要件を作成できる	
24	5	データ収集	★★★	クライアント技術	取得するデータが増えることを想定し、拡張性を考慮したクライアントアプリケーション（ロガーなど）を設計できる	
25	6	データ収集	★★★	クライアント技術	プラットフォームの違いを吸収し、統一的にデータを取得するプログラムを実装できる	
26	7	データ収集	★	通信技術	対象プラットフォームに用意された機能（HTTP、FTP、SSHなど）を用い、データを収集先に格納するための機能を実装できる	
27	8	データ収集	★★	通信技術	ネットワークプロトコルや暗号化などの通信技術を用い、通信のボトルネックを考慮した上で、必要な通信機能を実装できる	
28	9	データ収集	★★★	通信技術	データ通信において、機能・性能問題に対し根本原因を特定できるだけでなく、必要に応じて新規技術の適用を検討できる	
29	10	データ収集	★	データ統合	同種のデータを統合するシステムを設計できる	○
30	11	データ収集	★★★	データ統合	異種フォーマットが混在するデータを統合するシステムを設計できる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データエンジニアリングカ:チェック項目	必須スキル
31	12	データ収集	★★★	データ統合	システム分析・業務分析を元に、必要なデータフロー管理やジョブ管理ツールを選定・評価できる	
32	1	データ構造	★	基礎知識	扱うデータが、構造化データ（顧客データ、商品データ、在庫データなど）なのか非構造化データ（雑多なテキスト、音声、画像、動画など）なのか判断できる	○
33	2	データ構造	★	基礎知識	ER図を読んでテーブル間のリレーションシップを理解できる	○
34	3	データ構造	★	要件定義	業務で使用するシステムのデータのライフサイクル（いつ、どんなデータが発生し、いつまで保持されているのかなど）を把握して、論理モデルを作成できる	○
35	4	データ構造	★★	要件定義	データ保持ルール（データアクセス、性能、保持期間、セキュリティなど）に基づき、データベース・DWHの運用ルールを定義できる	
36	5	データ構造	★	テーブル定義	正規化手法（第一正規化～第三正規化）を用いてテーブルを正規化できる	
37	6	データ構造	★★	テーブル定義	ビジネスプロセスを理解・整理して、データフロー図、論理データモデル、ER図、テーブル定義書を作成できる	○
38	7	データ構造	★★	テーブル定義	業務特性や基幹システムの特徴を元に、検索で頻繁に使用するデータのキー（顧客IDなど）を想定し、インデックスを作成・設定できる	
39	8	データ構造	★★	テーブル定義	データ集計を高速化またはSQLを単純化するため、スタースキーマ、スノーflakeスキーマなどを用いたデータモデルを設計できる	
40	9	データ構造	★★	テーブル設計	稼働中の複数のシステム間で発生するデータ項目の差異を、変換テーブルを活用して、埋めることができる	
41	10	データ構造	★★	テーブル設計	DWHに入れる元データ（基幹DBのデータなど）のキーに変更があった場合に、サロゲートキーやナビゲーションブリッジテーブルを用いて対応できる	
42	11	データ構造	★★★	テーブル設計	非正規化テーブルや一時テーブルなどを作成し、アプリケーションの処理速度を高速化できる	
43	1	データ蓄積	★	DWH	DWHアプライアンス（Oracle Exadata、IBM Pure Data/Netezza、Teradataなど）に接続し、複数テーブルを結合したデータを抽出できる	
44	2	データ蓄積	★★	DWH	DWHアプライアンス（Oracle Exadata、IBM Pure Data/Netezza、Teradataなど）の機能と特徴を理解し、適切な管理対象データを選定できる	
45	3	データ蓄積	★	分散技術	Hadoopに対応可能なインターフェース（MapReduce、Pig、Hiveなど）を用いてHadoop上の大量データから条件に合致するデータを抽出できる	
46	4	データ蓄積	★★	分散技術	Hadoopの得意な点、苦手な点を理解し、Hadoopにて管理すべきデータを選定できる	○
47	5	データ蓄積	★★	分散技術	KVSの特性（集計・ソートが苦手、データの一意性保証など）を理解し、KVSがデータストア要件を満たすかを判断できる	
48	6	データ蓄積	★★★	分散技術	分散クラスタ構成が構築可能なRDBMS製品（Oracle RAC、IBM Pure Scaleなど）を用いてスケールアウト可能なオンプレミス構成を設計できる	
49	7	データ蓄積	★★★	分散技術	Hadoopの分散アーキテクチャを理解し、大容量データ処理のパフォーマンスチューニングができる	
50	8	データ蓄積	★	分散技術	NoSQLデータストア（HBase、Cassandra、Amazon DynamoDB、Couchbase、Azure DocumentDBなど）にAPIを介してアクセスし、新規データを登録できる	
51	9	データ蓄積	★	クラウド	クラウド上のストレージサービス（Amazon S3、Google Cloud Storage、Softlayer Object Storageなど）に接続しデータを格納できる	
52	10	データ蓄積	★★	クラウド	クラウド上のDWHサービス（Amazon Redshift、Google BigQueryなど）にデータをロードし公開できる	
53	11	データ蓄積	★★★	クラウド	クラウド上のデータストアサービスが機能面・非機能面で対象業務に合致するかの評価を行い、採用可否を判断できる	
54	12	データ蓄積	★★★	キャッシュ技術	基盤設計において、どこのシステム要素にmemcachedなどのキャッシュ機能を採用すると処理が高速化されるか判断できる	
55	13	データ蓄積	★★★	リアルタイムデータ分析	リアルタイムに入力されるストリームデータから指定条件のイベントを即時に抽出する複合イベント処理（CEP）を実現するサーバー環境・構成を設計できる	
56	14	データ蓄積	★★★	リアルタイムデータ分析	リアルタイムに連続して入力されるストリームデータの加工・集計処理を行うにあたり、Storm等のリアルタイム分散フレームワーク適用の有効性を判断できる	
57	15	データ蓄積	★★★	グリッドコンピューティング	分散処理のフレームワーク（Spark、Tezなど）を用いてアプリケーションの計算処理を複数サーバーに分散させる並列処理システムを設計できる	
58	16	データ蓄積	★★★	新規技術	データストアの技術動向に注目し、リレーショナルDBだけでなく、グラフDB・時系列DBなどの新規技術の検証・評価ができる	
59	1	データ加工	★	フィルタリング処理	数十万レコードのデータに対して、条件を指定してフィルタリングできる（特定値に合致する・もしくは合致しないデータの抽出、特定範囲のデータの抽出、部分文字列の抽出など）	○
60	2	データ加工	★★	フィルタリング処理	正規表現を活用して条件に合致するデータを抽出できる（メールアドレスの書式を満たしているか判定をするなど）	
61	3	データ加工	★	ソート処理	数十万レコードのデータに対して、レコード間で特定カラムでのソートができる。また、数千レコードのデータに対して、カラム間でソートできる	○

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データエンジニアリング力:チェック項目	必須スキル
62	4	データ加工	★	結合処理	数十万レコードのデータに対して、単一条件による内部結合、外部結合、自己結合ができる。また、UNION処理ができる	
63	5	データ加工	★	クレンジング処理	数十万レコードのデータに対して、NULL値や想定外・範囲外のデータを持つレコードを取り除く、または既定値に変換できる	○
64	6	データ加工	★★	クレンジング処理	フラットファイルやバイナリファイルに対するデータロードの前処理（クレンジング操作、禁則処理やバイナリ処理）ができる	
65	7	データ加工	★★	クレンジング処理	線形補間など、複数のレコードを考慮したクレンジング処理ができる	
66	8	データ加工	★	マッピング処理	数十万レコードのデータに対して、規定されたリストと照合して変換する、都道府県名からジオコードに変換するなど、ある値を規定の別の値で表現できる	
67	9	データ加工	★★★	マッピング処理	データ定義や実際の観測データの状況をもとに、名寄せ処理を設計・実装できる	
68	10	データ加工	★	サンプリング処理	数十万レコードのデータに対して、ランダムまたは一定間隔にデータを抽出できる	
69	11	データ加工	★	集計処理	数十万レコードのデータを集計して、合計や最大値、最小値、レコード数を算出できる	○
70	12	データ加工	★	変換・演算処理	数十万レコードのデータに対する四則演算ができる。また、数値データを日時データに変換するなど別のデータ型に変換できる	
71	13	データ加工	★★	変換・演算処理	数千万レコードのデータに対して、カラムナー型のデータに変換できる	
72	1	データ共有	★	データ出力	加工・分析処理結果をCSV、XML、Excelなどの指定フォーマット形式に変換してエクスポートできる	○
73	2	データ共有	★	データ出力	加工・分析処理結果を、接続先DBのテーブル仕様に合わせてレコード挿入できる	○
74	3	データ共有	★	データ展開	データ取得用のWeb API (REST) やWebサービス (SOAP) などを用いて、必要なデータを取得できる	
75	4	データ共有	★★	データ展開	利用者の要件に合致したレポート (図、表) を、PDFやPostScriptなどの印刷用フォーマットで出力する変換機能を設計できる	
76	5	データ共有	★★	データ展開	BIツール (Tableau、QlikView、PowerBI、Cognos、SASなど) 用のサーバーを構築し、単一データソースのレポート用DBを設計・構築できる	
77	6	データ共有	★★	データ展開	データストア上のデータをメールやメッセージ (Webサービスなど) を用いてプッシュ配信するシステムのサーバー・ネットワーク・ソフトウェアの構成を設計できる	
78	7	データ共有	★★	データ展開	RSS、RDFや業界標準フォーマットなど要件に合致したデータ形式・配信形式で、情報提供するシステムのインターフェースを設計できる	
79	8	データ共有	★★	データ展開	Web API (REST) やWebサービス (SOAP) などを用いて、必要なデータを提供するシステムの公開インターフェースを設計できる	
80	9	データ共有	★	データ連携	FTPサーバー、ファイル共有サーバーから必要なデータファイルをダウンロードして、Excelなどの表計算ソフトに取り込み活用できる	○
81	10	データ共有	★	データ連携	BIツール (Tableau、QlikView、PowerBI、Cognos、SASなど) のレポート編集機能を用いて新規レポートを公開できる	
82	11	データ共有	★	データ連携	BIツールの自由検索機能を活用し、必要なデータを抽出して、グラフを作成できる	
83	12	データ共有	★★	データ連携	連携対象システムの仕様に合わせて、ETLツールを用いたデータ変換、ファイル転送処理を実装できる	
84	13	データ共有	★★★	データ連携	ESB・EAIなどのデータ連携基盤を活用してシステム間のデータ連携 (データ配信・交換) を行いインターフェースを設計できる	
85	1	プログラミング	★	基礎プログラミング	小規模な構造化データ (CSV、RDBなど) を扱うデータ処理 (抽出・加工・分析など) を、設計書に基き、プログラム実装できる	○
86	2	プログラミング	★	データインタフェース	JSON、XMLなど標準的なフォーマットのデータを受け渡すために、APIを使用したプログラムを設計・実装できる	
87	3	プログラミング	★★	アルゴリズム	非効率なループ処理の性能向上などのために、計算量やメモリを意識したプログラム実装ができる	
88	4	プログラミング	★★★	アルゴリズム	最新の論文に発表された数式処理や機械学習ロジックをプログラム実装できる	
89	5	プログラミング	★★	拡張プログラミング	データ型を意識したプログラム実装・試験ができる (C言語での性能・誤差を意識したデータ型の実装、Pythonでのエラー時のデータ型を意識したデバッグなど)	
90	6	プログラミング	★★	拡張プログラミング	GPU (GPGPU) 環境を処理するライブラリを有効に活用できる	
91	7	プログラミング	★★	拡張プログラミング	異なるタイプの複数の処理を効率よく行うために、スクリプトを用いたプログラムを設計・実装できる (パイプライン処理のluigiなど)	
92	8	プログラミング	★★★	拡張プログラミング	対象プラットフォーム (iOS、Android、HEMSなど) におけるデータ取得の業界標準を理解しており、今後の技術動向や規制についてのリスクを提示できる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	データエンジニアリングカ:チェック項目	必須スキル
93	9	プログラミング	★★	データ規模	SNSから抽出した非構造化データを、適切な言語やライブラリを選んでプログラム実装できる	
94	10	プログラミング	★★★	データ規模	単一サーバーの物理メモリを超える複数のデータソースを組合わせたデータ処理において、分散処理アーキテクチャやデータのインメモリ処理の特性を意識してプログラム設計ができる	
95	11	プログラミング	★★	分析プログラム	分析プログラムのロジックと処理手順を理解した上で正しい分析結果を出力しているか検証ができる	○
96	12	プログラミング	★★★	リアルタイム処理	ストリーミング処理や複合イベント処理（CEP）などを設計し、適切な言語やライブラリを選んでプログラム実装できる	
97	13	プログラミング	★	SQL	SQLで簡単なSELECT文を記述・実行できる（副問合せを含まない、2テーブル程度の結合と簡単なWHERE条件を含むSELECT文）	○
98	14	プログラミング	★★	SQL	SQLの構文を一通り知っていて、記述・実行できる（DML・DDLの理解、各種JOINの使い分け、集計関数とGROUP BY、CASE文を使用した縦横変換、副問合せやEXISTSの活用など）	○
99	15	プログラミング	★★★	SQL	RDBにおける分析関数の構文と挙動を理解し、分析関数を用いて複雑な副問合せや自己結合を解消できる	
100	16	プログラミング	★★★	SQL	N:Nの結合や完全外部結合の危険性（計算量の増大、結果の不完全性）、暗黙の型変換の危険性（インデックス不使用、小数点以下の切り捨てなど）を考慮したSQLを記述できる	
101	17	プログラミング	★★★	SQL	記述したSQLの実行計画の確認と判断ができ、SQLの修正やインデックス作成により、処理時間を大幅に改善するようなパフォーマンスチューニングができる	
102	18	プログラミング	★★	Pig/Hive	Hiveを記述して、パーティションが切られているデータを適切に処理できる	
103	19	プログラミング	★★	Pig/Hive	Pigを記述して列数・内容の異なる行が混在しているデータセットやネスト構造を持つデータセットを処理できる	
104	20	プログラミング	★★★	Pig/Hive	HiveまたはPigで使用するためのUDFが実装できる	
105	1	ITセキュリティ	★	基礎知識	セキュリティの3要素（機密性、可用性、完全性）について具体的な事例を用いて説明できる	○
106	2	ITセキュリティ	★★	プライバシー	データ匿名化の方法を理解し、匿名化方法（ハッシュ化、マスキング、k-匿名化など）に応じた加工処理を設計できる	○
107	3	ITセキュリティ	★	攻撃と防御手法	コンピュータウイルスとワームの違いを説明でき、データやシステムを防御する方法を理解している	
108	4	ITセキュリティ	★	攻撃と防御手法	ファイルシステム、ネットワーク、アプリケーション、データに対するユーザーごとのアクセスレベルを手順に従い設定できる	
109	5	ITセキュリティ	★	攻撃と防御手法	DoS攻撃、不正アクセス、マルウェアなど外部からの攻撃が発覚した場合に既存のルールに基づき対応できる	
110	6	ITセキュリティ	★★	攻撃と防御手法	ファイルシステム、ネットワーク、アプリケーション、データに対するユーザーごとのアクセスレベルを設計できる	
111	7	ITセキュリティ	★★	攻撃と防御手法	SQLインジェクションやバッファオーバーフローなどの素通りを防止する対策を判断できる	
112	8	ITセキュリティ	★★★	攻撃と防御手法	なりすまし、改ざん、中間者攻撃による盗聴などを防御するための対策を特定できる	
113	9	ITセキュリティ	★★★	攻撃と防御手法	侵入検知システム（IDS）やファイアウォールなど外部からの不正アクセスを検知、防御する環境を設計できる	
114	10	ITセキュリティ	★★★	攻撃と防御手法	個別の案件ごとに、依頼元との契約約款、依頼元がデータをどのように保持し利用するかに応じて、適切な匿名化の手法を選択し適用できる	
115	11	ITセキュリティ	★	暗号化技術	暗号化されていないデータは盗聴される恐れがあることを理解し、データの機密度合いに応じてソフトウェアを使用した暗号化と復号化ができる	○
116	12	ITセキュリティ	★	暗号化技術	なりすましや偽造された文書でないことを証明するために電子署名が必要であることを理解している	
117	13	ITセキュリティ	★	暗号化技術	一方方向ハッシュ関数を用いて、データの改ざんを検出できる	
118	14	ITセキュリティ	★★	暗号化技術	SSHやSSL/TLSなどのセキュアプロトコルの概要と必要性を説明できる	
119	15	ITセキュリティ	★★★	暗号化技術	DES、AES、RC4を用いたKerberos認証が使われる事例と仕組みを説明できる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	ビジネスカ：チェック項目	必須スキル
1	1	行動規範	★	ビジネスマインド	ビジネスにおける論理とデータの重要性を認識し、分析的でデータドリブンな考え方に基づき行動できる	○
2	2	行動規範	★★	ビジネスマインド	ビジネスではスピード感がより重要であることを認識し、時間と情報が限られた状況下でも、言わば「ザックリ感」を持って素早く意思決定を行うことができる	
3	3	行動規範	★★	ビジネスマインド	作業ありきではなく、本質的な問題（イシュー）ありきで行動できる	○
4	4	行動規範	★★	ビジネスマインド	最終的な結論に関わる部分や、ストーリーラインの骨格に大きな影響を持つ部分から着手するなど、取り組むべき分析上のタスクの優先度を判断できる	
5	5	行動規範	★★★	ビジネスマインド	プロフェッショナルとして、作業量ではなく、生み出す価値視点で常に判断、行動でき、依頼元にとって真に価値あるアウトプットを生み出すことをコミットできる	○
6	6	行動規範	★	データ倫理	データを取り扱う人間として相応しい倫理を身に着けている（データのねつ造、改ざん、盗用を行わないなど）	○
7	7	行動規範	★★	データ倫理	チーム全員がデータを取り扱う人間として相応しい倫理を持てるよう、適切にチームを管理できる	
8	8	行動規範	★★★	データ倫理	データの取り扱いに関する、会社や組織全体の倫理を維持、向上させるために、必要な制度や仕組みを策定し、その運営を主導することができる	
9	9	行動規範	★	法令	個人情報に関する法令の概要を理解している	○
10	10	行動規範	★★	法令	担当するビジネスや業界に関係する法令を理解しており、データの保持期間や運用ルールに活かすことができる	○
11	11	行動規範	★★	法令	個人情報の扱いに関する法令、その他のプライバシーの問題、依頼元との契約約款に基づき、匿名化するべきデータを選別できる（名寄せにより個人を特定できるもの、依頼元がデータ処理の結果をどのように保持し利用するのかなども考慮して）	
12	1	論理的思考	★	MECE	データや事象の重複に気づくことができる	○
13	2	論理的思考	★★	MECE	初見の領域に対して、抜け漏れや重複をなくすことができる	
14	3	論理的思考	★★★	MECE	未知の領域であっても、類似する事象の推測などを活用し、抜け漏れや重複をなくすことができる	
15	4	論理的思考	★★	データをもとに論理的に捉える	様々なデータや事象を、階層やグルーピングによって、構造化できる（ピラミッド構造）	
16	5	論理的思考	★★	データをもとに論理的に捉える	仮説思考を用いて、論点毎に分析すべき点を識別できる	
17	6	論理的思考	★★★	データをもとに論理的に捉える	論理的な整理にとらわれず、批判的・複合的な視点で課題を識別できる	
18	7	論理的思考	★	言語化能力	分析結果の意味合いを適切に言語化できる	○
19	8	論理的思考	★★	言語化能力	近い概念がある場合、データ表現に対して適切な言葉を選択できる	
20	9	論理的思考	★★★	言語化能力	データ表現に適した言葉がない場合でも、適切な言葉を新たに作り出すことができる	
21	10	論理的思考	★	ストーリーライン	一般的な論文構成について理解している（序論⇒アプローチ⇒検討結果⇒考察や、序論⇒本論⇒結論 など）	
22	11	論理的思考	★★	ストーリーライン	因果関係に基づいて、ストーリーラインを作る（観察⇒気づき⇒打ち手、So What?, Why So?など）	
23	12	論理的思考	★★★	ストーリーライン	相手や内容に応じて、自在にストーリーラインを組み上げることができる	
24	13	論理的思考	★	ドキュメンテーション	1つの図表～数枚程度のドキュメントを論理立ててまとめることができる（課題背景、アプローチ、検討結果、意味合い、ネクストステップ）	
25	14	論理的思考	★★	ドキュメンテーション	10～20枚程度のミニパッケージ（テキスト&図表）、もしくは5ページ程度の図表込みのビジネスレポートを論理立てて作成できる	
26	15	論理的思考	★★★	ドキュメンテーション	30～50枚程度のフルパッケージ（テキスト&図表）、もしくは10ページ以上のビジネスレポートを論理立てて作成できる	
27	16	論理的思考	★	プレゼンテーション	プレゼンテーションの論拠不足や論理破綻を指摘された際に、すみやかに理解できる	
28	17	論理的思考	★★	プレゼンテーション	論理的なプレゼンテーションができる	
29	18	論理的思考	★★★	プレゼンテーション	プレゼンテーションの相手からの質問や反論に対して、説得力のある形で回答できる	
30	1	プロセス	★	基礎知識	目的やゴールの設定がないままデータを分析しても、意味合いが出ないことを理解している	○

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	ビジネス力：チェック項目	必須スキル
31	2	プロセス	★	基礎知識	課題や仮説を言語化することの重要性を理解している	
32	3	プロセス	★	KPI	一般的な収益方程式、バランスシートなどに加え、自らが担当する業務の主要な変数（KPI）を理解している	
33	4	プロセス	★★	KPI	自らが関連する事業領域であれば、複数の課題レイヤーにまたがっていても、KPIを整理・構造化できる	
34	5	プロセス	★★★	KPI	初見の事業領域であっても、KPIを構造化し、重要なKPIを見極められる	
35	6	プロセス	★	問題の大枠定義	自分の担当する業界について、市場規模、主要なプレーヤー、支配的なビジネスモデル、課題と機会について説明できる	○
36	7	プロセス	★★	問題の大枠定義	事業モデルやバリューチェーンなどの特徴や事業の主たる課題を自力で構造的に理解でき、問題の大枠を整理できる	
37	8	プロセス	★★★	問題の大枠定義	領域の主要課題を他領域の課題との関連も含めて構造的に理解でき、問題の大枠を定義できる	
38	9	プロセス	★	課題の定義	担当する業務領域であれば、基本的な課題の枠組みが理解できる（調達活動を5フォースで整理する、CRM課題をRFMで整理するなど）	
39	10	プロセス	★★	課題の定義	取り扱う課題領域に対して、新規課題の切り分けや枠組み化ができる	
40	11	プロセス	★★★	課題の定義	仮説や可視化された問題がなくとも、解くべき課題を構造的に整理でき、見極めるべき論点を特定できる	○
41	12	プロセス	★	プロジェクト推進	ウォーターフォール開発とアジャイル開発の違いを理解している	
42	13	プロセス	★★	プロジェクト推進	アジャイル開発体制のポイントを理解した上で、アジャイルな開発チームを立ち上げ、推進することができる	
43	14	プロセス	★	アプローチ設計	スコープ、検討範囲・内容が明快に設定されていれば、必要なデータ、分析手法、可視化などを適切に選択できる	
44	15	プロセス	★★	アプローチ設計	解くべき課題がフレーミングされていれば、必要なデータ、分析手法、可視化などを適切に選択できる	○
45	16	プロセス	★★★	アプローチ設計	複数の事業や課題にまたがっていても、必要なデータ、分析手法、可視化などを適切に選択し作業手順に落とし込める	
46	17	プロセス	★★★	アプローチ設計	複数のアプローチの組み合わせでしか解けない課題であっても、その解決までの道筋を設計できる	
47	18	プロセス	★★★	アプローチ設計	他社による模倣を防ぐなど、競争力を保つ観点でアプローチの設計ができる（機械化や標準化の範囲を絞るなど）	
48	19	プロセス	★★	分析価値の判断	分析で解くべき課題が否かを判断できる	
49	20	プロセス	★★★	分析価値の判断	プロジェクトの開始時点で、入手可能なデータ、分析手法、インフラ、ツールの生み出すビジネス価値を適切に見積りができる	
50	1	データの理解・検証	★	統計情報への感度	ニュース記事などで統計情報に接したときに、数字やグラフの持つメッセージを理解できる	
51	2	データの理解・検証	★★	統計情報への感度	積極的に統計情報を収集しているとともに、表現に惑わされず数字を正当に評価できる（原点が0ではないグラフ、不当に誇張されたグラフなど）	
52	3	データの理解・検証	★★★	統計情報への感度	数字やデータの検証のために、何と比較するべきかすみやかに把握し、収集・利用できる（業務データや過去に接触した統計情報の想起・活用を含む）	
53	4	データの理解・検証	★	データ確認	単独のグラフに対して、集計ミスなどがないかチェックできる	○
54	5	データの理解・検証	★	データ確認	データ項目やデータの量・質について、指示のもと正しく検証し、結果を説明できる	
55	6	データの理解・検証	★★	データ確認	複数のグラフや集計表で構成されているレポートに対して、全体として集計ミスや不整合が起きていないかチェックできる	○
56	7	データの理解・検証	★★	データ確認	データ項目やデータの量・質の検証方法を計画・実行し、その結果をもとにその後の分析プロセスを立案・修正できる	
57	8	データの理解・検証	★★★	データ確認	多数のグラフ、集計表、外部の統計情報、高度なデータ解析手法を用いた解析結果などを含むレポートに対して、不整合が起きていないか、妥当性の高い論理構造であるかチェックできる	
58	9	データの理解・検証	★★★	データ確認	分析に必要なデータを想定し、現在取得可能なデータで量・質ともに分析に耐える内容であるか判断できる。また、目的が達成可能であるか見込みが立てられる	○
59	10	データの理解・検証	★	俯瞰・メタ思考	データの裏を読むことの重要性を理解している	
60	11	データの理解・検証	★★	俯瞰・メタ思考	データを俯瞰して、変化をすみやかに察知できる。また、変化が誤差の範囲かどうか判断できる	○
61	12	データの理解・検証	★★★	俯瞰・メタ思考	複数のデータを多元的かつ大局的に俯瞰して、大きな動きや本質的な事実を見抜くことができる	○
62	13	データの理解・検証	★	データ理解	データから事実を正しく浮き彫りにするために、集計の切り口や比較対象の設定が重要であることを理解している	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	ビジネススキル: チェック項目	必須スキル
63	14	データの理解・検証	★	データ理解	普段業務で扱っているデータの発生トリガー・タイミング・頻度などを説明でき、また基本統計量を把握している	
64	15	データの理解・検証	★	データ理解	何のために集計しているか、どのような知見を得たいか、目的に即して集計できる	○
65	16	データの理解・検証	★★	データ理解	生データを眺めて、どのような切り口で集計・比較すればデータの理解や事実の把握につながるか検討できる	○
66	17	データの理解・検証	★★	データ理解	扱ったことのない新たなデータに内容の不明な項目があっても、生データの閲覧や集計を通して何の項目かあたりをつけられる	
67	18	データの理解・検証	★★	データ理解	扱っているデータの関連業務の知識と分析目的を踏まえて、どんな説明変数が効きそうか、あたりをつけて洗い出し、構造的に整理できる	
68	19	データの理解・検証	★★★	データ理解	データの変化の裏で起きている事実を読み取り、仮説を立て検証方法を立案できる	
69	20	データの理解・検証	★★★	データ理解	データを入手する前に、存在するであろうデータとその分布を想定して基礎俯瞰の方向性やその結果の想定ができ、それを前提とした解析方法の検討・ラフ設計をすることができる	
70	21	データの理解・検証	★★★	データ理解	扱ったことのない新たなデータであっても、ER図やテーブル定義、生データなどを見ることによってデータの発生源や欠損値の意味などのあたりをつけられる	
71	22	データの理解・検証	★★	データ粒度	分析目的とデータの量・質を踏まえて、想定されるメッセージと統計的観点から適切な集計単位とサンプリング率を決定できる	
72	23	データの理解・検証	★★★	データ粒度	分析目的とデータの量・質に加えて、想定しているメッセージ、深掘りの方向性・可能性、処理負荷、データ処理フローなども総合的に踏まえた最適な集計単位とサンプリング率を決定できる	
73	24	データの理解・検証	★	ビジネス観点	ビジネス観点で仮説を持ってデータをみることの重要性和、仮説と異なる結果となったときにそれが重大な知見である可能性を理解している	
74	25	データの理解・検証	★★	ビジネス観点	特定の分析プロセスにおいて、統計手法による閾値や係数（量的変数の統計的カテゴリズなど）に対して商品単価や購入周期を考慮するなど、ビジネス観点で納得感のある調整ができる	
75	26	データの理解・検証	★★★	ビジネス観点	分析プロセス全体を通して常時、ビジネス観点での妥当性をチェックしている。また、データから得られた示唆がビジネス観点で新しい知見であるか都度判断できる	○
76	1	データ入手	★	データ入手	仮説や既知の問題が与えられた中で、必要なデータにあたりをつけ、入手できる	○
77	2	データ入手	★	データ入手	現場に向いてヒアリングするなど、一次情報に接することの重要性を理解している	
78	3	データ入手	★★	データ入手	自身が担当するプロジェクトやサービスを越えて、必要なデータのあたりをつけることができる	○
79	4	データ入手	★★	データ入手	既存のサービスやアプリケーションに対して、有効な分析をするためのログ出力の追加仕様を整理することができる	
80	5	データ入手	★★★	データ入手	組織全体及び関連する社外のデータを見渡して、必要なデータのあたりをつけることができる	
81	6	データ入手	★★★	データ入手	入手可能なデータに加え、分析結果の品質・効果を向上させる新たなデータ入手方法を提案できる（IoTでの新設センサーの種類・配置場所・データ入手間隔など）	
82	1	意味合いの抽出、洞察	★★★	開示	担当および関連する分析プロジェクトのデータ、分析結果を顧客、外部に開示すべきか判断できる	
83	2	意味合いの抽出、洞察	★	洞察	分析、図表から直接的な意味合いを抽出できる (バラツキ、有意性、分布傾向、特異性、関連性、変曲点、関連度の高低など)	○
84	3	意味合いの抽出、洞察	★	洞察	想定に影響されず、分析結果の数値を客観的に解釈できる	
85	4	意味合いの抽出、洞察	★★	洞察	ビジネス上の意味を捉えるために、特異点、相違性、傾向性、関連性を見出したうえで、ドメイン知識を持つ人に適切な質問を投げかけられる	
86	5	意味合いの抽出、洞察	★★	洞察	データの特徴を見て意味合いの明確化に向けた分析の深掘り、データ見直しの方向性を設計できる	
87	6	意味合いの抽出、洞察	★★★	洞察	担当および関連する分析プロジェクトのデータを眺め、裏にある構造や意味合い（真実）を見ぬくことができる	
88	7	意味合いの抽出、洞察	★★	評価	担当する分析プロジェクトの分析結果を見て検討目的と合っているか再評価できる	○
89	8	意味合いの抽出、洞察	★★	評価	各種の解析手法（主成分分析、クラスター分析、決定木分析など）の結果を解釈し、意味合いを適切に表現・説明できる	
90	9	意味合いの抽出、洞察	★★	評価	分析結果が当初の目的を満たしていない場合に、必要な分析手順を追加できる	
91	10	意味合いの抽出、洞察	★★★	評価	分析結果が当初の目的を満たしていない場合に、必要に応じてプロジェクト全体を再設計できる	
92	1	解決	★★	総合的評価	分析内容を総合的に評価し、モデルの定性的な評価ができる	

No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	ビジネスカ:チェック項目	必須スキル
93	2	解決	★★	提案	分析的検討に基づき、担当業務に対する必要なアクション、改革案を整理して結論を導くことができる	
94	3	解決	★★★	提案	分析的検討に基づき、経営レベルで必要なアクション、改革案を整理して結論を導くことができる	○
95	4	解決	★★★	提案	データを利活用した持続性のある事業モデルを設計できる	
96	1	事業に実装する	★★	実装	現場に実装する際、現場での実行可能性を考慮し適切に対応できる	
97	2	事業に実装する	★★	実装	担当する案件が自分の持っている予算内で解決するように取り組みをデザインし、実装できる	
98	3	事業に実装する	★★★	実装	費用対効果、実行可能性、業務負荷を考慮した実装ができる	
99	4	事業に実装する	★★	異分野とのコミュニケーション	異なるスキル分野の専門家や事業者と適切なコミュニケーションをとりながら実装を進めることができる	○
100	5	事業に実装する	★	モニタリング	やりっぱなしではなく結果、改善の度合いをモニタリングする重要性を理解している	
101	6	事業に実装する	★★	モニタリング	事業・現場へ実装するにあたりモニタリングの仕組みを適切に組み込むことができる	
102	7	事業に実装する	★★★	モニタリング	既存のPDCAに新たな仕込みを行い、次の改善的な取り組みにつなげることができる	
103	8	事業に実装する	★★★	横展開	特定のビジネス課題に向けた新しいソリューションを個別の現場の特性を考慮し横展開できる	
104	1	活動マネジメント	★★★	データ倫理	ウイルス、DDoS攻撃などの深刻なセキュリティ攻撃を受けた場合に対応する最新の技術を熟知し、対応する専門組織（CSIRT）の構成を責任者にすみやかに提案できる	
105	2	活動マネジメント	★	リソースマネジメント	指示に従ってスケジュールを守り、チームリーダーに頼まれた自分の仕事を完遂できる	○
106	3	活動マネジメント	★★	リソースマネジメント	自身とチームメンバーのスキルを把握し、適切なプロジェクト管理ができる	○
107	4	活動マネジメント	★★	リソースマネジメント	担当するプロジェクトで、設定された予算やツール、システム環境を適切に活用し、プロジェクト進行できる	
108	5	活動マネジメント	★★	リソースマネジメント	5名前後のチームを設計し、スケジュール通りに進行させ、ステークホルダーに対して期待値に見合うアウトプットを安定的に生み出せる	
109	6	活動マネジメント	★★★	リソースマネジメント	プロジェクトに求められるスキル要件と各メンバーのスキル・成長目標・性格をふまえ、現実的にトレードオフ解消とシナジーを狙ったリソースマネジメントができる	
110	7	活動マネジメント	★★★	リソースマネジメント	プロジェクトメンバーの技量を把握した上で、プロジェクト完遂に必要なツール選定、予算策定、スコープ設定、またはアウトソーシング体制を検討・構築できる	
111	8	活動マネジメント	★★★	リソースマネジメント	複数のチームを設計し（総合的なプロジェクトマネジメント）、スケジュール通りに進行させ、複合的なステークホルダーに対し、期待値を超えたアウトプットを安定的に生み出せる	
112	9	活動マネジメント	★	リスクマネジメント	担当するタスクの遅延や障害などを発見した場合、プロジェクトオーナーに迅速かつ適切に報告ができる	
113	10	活動マネジメント	★★	リスクマネジメント	担当するプロジェクトでの遅延や障害などの発生を検知し、リカバリーするための提案・設計ができる	
114	11	活動マネジメント	★★★	リスクマネジメント	プロジェクトに何らかの遅延・障害などが発生した場合、適切なリカバリー手順の判断、リカバリー体制構築、プロジェクトオーナーに対する迅速な対応ができる	○
115	12	活動マネジメント	★★	育成/ナレッジ共有	自身とチームメンバーのスキルを大まかに把握し、担当するプロジェクトを通してチームメンバーへのスキルアドバイスやスキル成長のための目標管理ができる	
116	13	活動マネジメント	★★	育成/ナレッジ共有	チームメンバーのスキルに応じ、研修参加や情報収集への適切なアドバイスやチーム内でのナレッジ共有を推進できる	
117	14	活動マネジメント	★★★	育成/ナレッジ共有	チームの各メンバーに対し、データサイエンティストとしてのスキル目標の設定、到達させるためのプロジェクトを通じた適切なアドバイスができる	
118	15	活動マネジメント	★★★	育成/ナレッジ共有	チーム育成の上で、データサイエンティストに求められるスキルについて、研修制度の設計やナレッジ共有の仕組み構築と運営ができる	
119	16	活動マネジメント	★★★	育成/ナレッジ共有	チームに必要な情報やデータサイエンスの新しい技術・手法に関する情報収集戦略やラーニング方法を主導し、自ら情報を取捨選択し、チームにフィードバックできる	
120	17	活動マネジメント	★★	営業マネジメント	ビジネス要件を整理し、分析・データ活用のプロジェクトを企画・提案することができる	
121	18	活動マネジメント	★★★	営業マネジメント	依頼元やステークホルダーのビジネスをデータ面から理解し、分析・データ活用のプロジェクトを立ち上げ、プロジェクトにかかるコストと依頼元の利益を説明できる	
122	19	活動マネジメント	★★	組織マネジメント	データサイエンスチームの役割を認識し、担当するプロジェクトにおいて、組織内や他部門・他社間でのタスク設定や調整ができる	



No	Sub No	スキルカテゴリ	スキルレベル	サブカテゴリ	ビジネスカ:チェック項目	必須スキル
123	20	活動マネジメント	★★★	組織マネジメント	データサイエンスチームを自社・他社の様々な組織と関連付け、組織内での役割の規定、組織的な目標設定を行う	

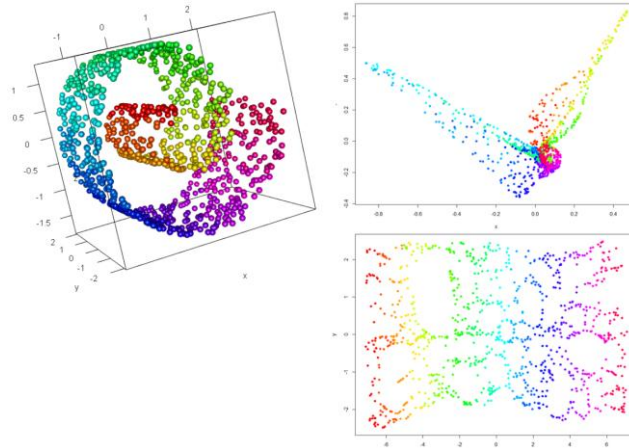
【添付別紙】

スキルチェックリスト：データサイエンスカ

Data visualizationに関するイメージ共有のための参考資料（例示）

項目NO

10 データの持つ分散量の観点で、高次のデータを主成分分析(PCA)などにより1~3次元のデータに変換できる

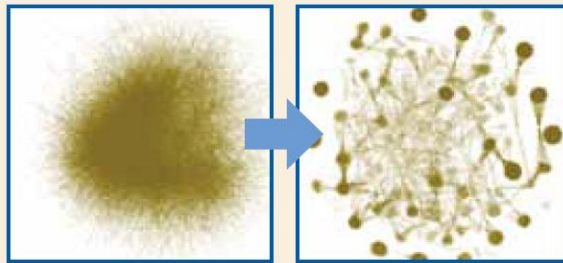


引用元：株式会社ALBERT WEBサイト  
[http://www.albert2005.co.jp/analyst\\_blog/?p=1214](http://www.albert2005.co.jp/analyst_blog/?p=1214)

13 ネットワーク構造、グラフ構造などの表現において、ノードとエッジが増えすぎて特徴抽出が困難であっても、データの絞り込みや抽象度を上げることで適切に可視化できる

図2 企業間取引データをクラスター化して把握しやすくした例

	研磨前	研磨後
頂点数	3,282	3,282
枝数	35,168	73,132
クリーク数	32,953	343



引用元：（出版物）大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 発行  
「国立情報学研究所ニュース [NII Today] 第66号 平成26年12月」 6ページ  
[http://www.nii.ac.jp/userdata/results/pr\\_data/NII\\_Today/66/all.pdf](http://www.nii.ac.jp/userdata/results/pr_data/NII_Today/66/all.pdf)

19 強調表現がもたらす効果を理解している(計量データに対しては位置やサイズ表現が色表現よりも効果的など)

表現例	量を表現する	順序を表現する	分類を表現する	関係を表現する
位置 	✓	✓	✓	✓
長さ 	✓	✓		
明度/彩度・色 		✓		✓

引用元：「solutions-2」WEBサイト  
<http://www.solutions2.be/Books.aspx> をもとに独自に修正加筆

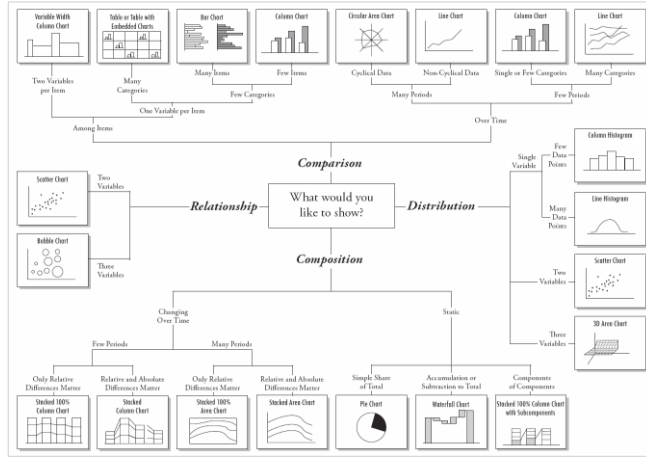
【添付別紙】

スキルチェックリスト：データサイエンスカ

Data visualizationに関するイメージ共有のための参考資料（例示）

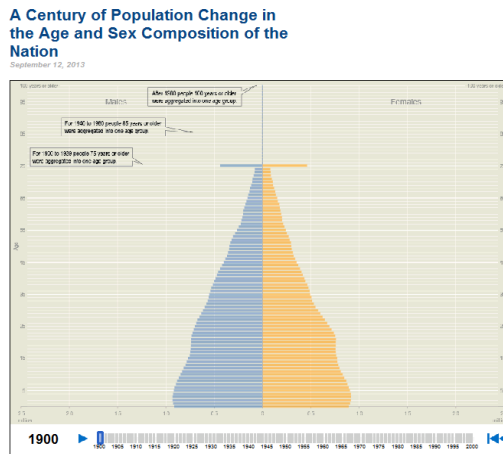
項目NO

20 1～3次元の比較において目的(比較、構成、分布、変化など)に応じた図表化ができる



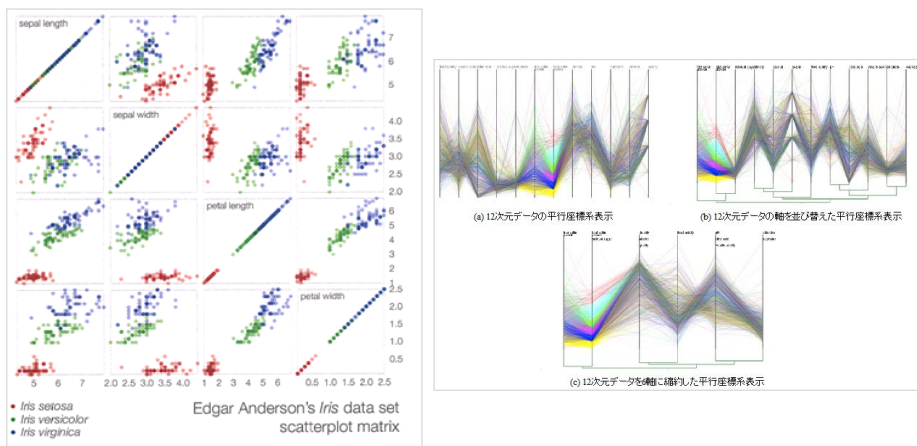
引用元： Amit Agarwal 「Digital Inspiration」 WEBサイト  
<http://www.labnol.org/software/find-right-chart-type-for-your-data/6523/>

21 端的に図表の変化をアニメーションで可視化できる(人口動態のヒストグラムが経年変化する様子を表現するなど)



引用元： アメリカ国税調査局 WEBサイト  
<http://www.census.gov/dataviz/visualizations/055/>

23 1～3次元の図表を拡張した多変量の比較(平行座標、散点図行列、テーブルレンズ、ヒートマップなど)を適切に可視化できる



引用元： [左図] 「Data-Driven Documents」 WEBサイト <http://mbostock.github.io/d3/talk/20111116/iris-splom.html>  
 [右図] 「グラスペクトル解析を用いた軸縮約可能平行座標系」 WEBサイト  
<http://user.keio.ac.jp/~yun/projects/cpc/index-j.html>

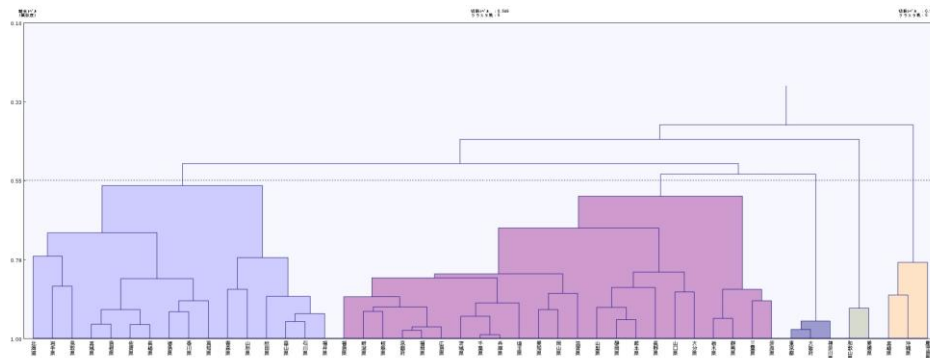
【添付別紙】

スキルチェックリスト：データサイエンティスト

Data visualizationに関するイメージ共有のための参考資料（例示）

項目NO

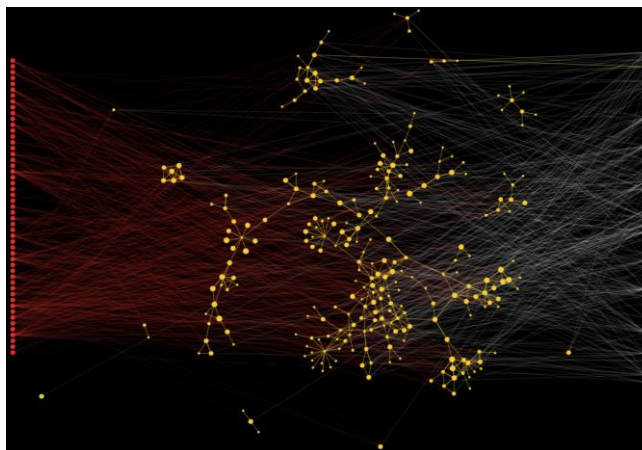
24 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



引用元：株式会社ALBERT WEBサイト

[http://www.albert2005.co.jp/technology/images/tech\\_mining\\_img106.jpg](http://www.albert2005.co.jp/technology/images/tech_mining_img106.jpg)

24 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



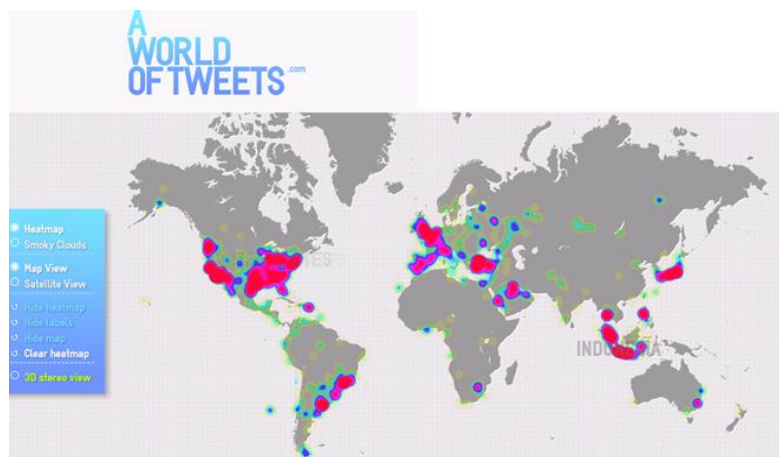
赤・白点：Wine

黄点：Cheese

引用元：「Cytoscape.js」WEBサイト（Wine & cheese）

<http://cytoscape.github.io/cytoscape.js/demos/cde4db55e581d10405f5/>

25 GPSデータなどを平面地図上に重ね合わせた可視化ができる



引用元：「A WORLD OF TWEETS」WEBサイト

<http://aworldoftweets.frogdesign.com/>

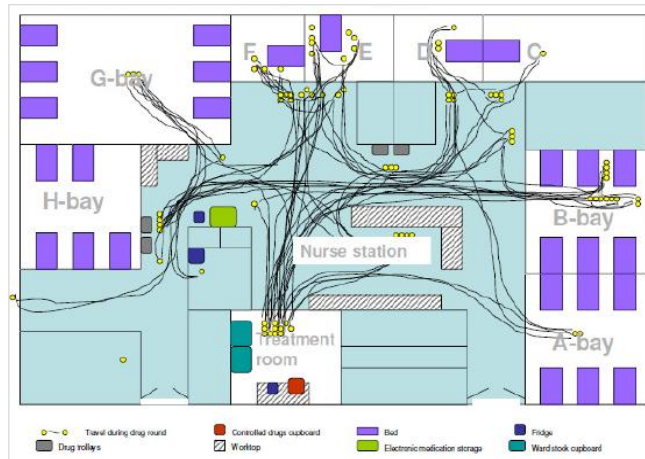
【添付別紙】

スキルチェックリスト：データサイエンティスト

Data visualizationに関するイメージ共有のための参考資料（例示）

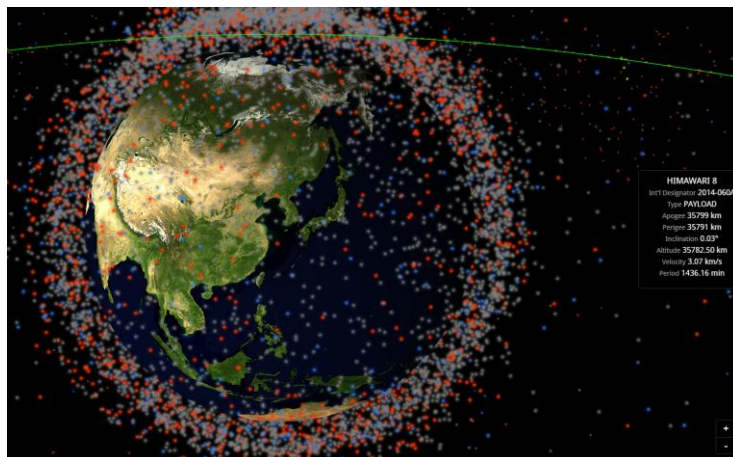
項目NO

26 挙動・軌跡の可視化ができる(店舗内でのユーザの回遊やEye trackingなど)



引用元：「Imperial College London」WEBサイト  
<https://www1.imperial.ac.uk/cpssq/ms/m/>

28 人体、標高を持つ地図、球面などの上にデータを重ね合わせた可視化ができる



引用元：James Yoder「Stuff in Space」WEBサイト  
<http://stuffin.space/>

29 地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる(多地点での風の動き、飛行物の軌跡など)



引用元：Kaspersky Lab「CYBERTHREAT REAL-TIME MAP」WEBサイト  
<https://cybermap.kaspersky.com/>

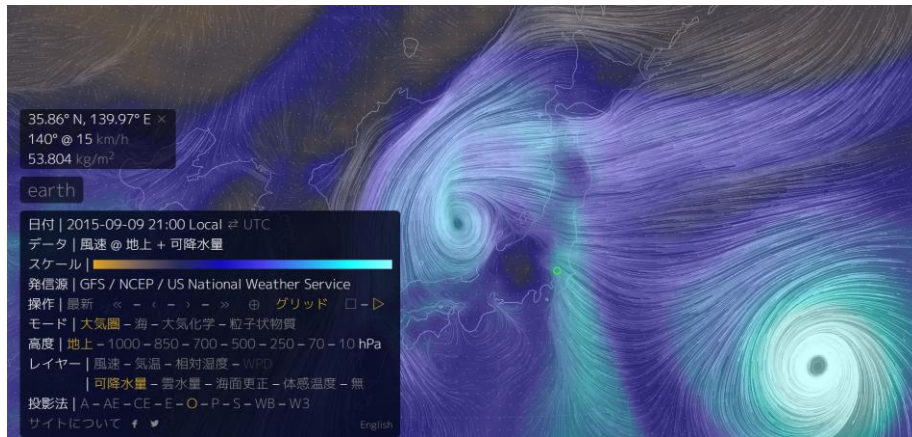
【添付別紙】

スキルチェックリスト：データサイエンシカ

Data visualizationに関するイメージ共有のための参考資料（例示）

項目NO

29 地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる(多地点での風の動き、飛行物の軌跡など)



引用元：Cameron Beccario 「earth」 WEBサイト

[http://earth.nullschool.net/jp/#2015/09/09/1200Z/wind/surface/level/overlay=total\\_precipitable\\_water/orthographic=-224.70,38.05,3000/grid=on](http://earth.nullschool.net/jp/#2015/09/09/1200Z/wind/surface/level/overlay=total_precipitable_water/orthographic=-224.70,38.05,3000/grid=on)



# データ社会に求められる 新しい才能とスキル

第2回シンポジウム

2015年11月13日

一般社団法人 データサイエンティスト協会

スキル委員会

## 設立時の状況

- バズワードである「データサイエンティスト」という言葉の定義が欠落
- 雇い主側の期待に雇われる側のスキルセットが合致しないケースが増加
- 若い才能たちが、自分達をどう訓練し、スキルを身に付けていくべきかわからない

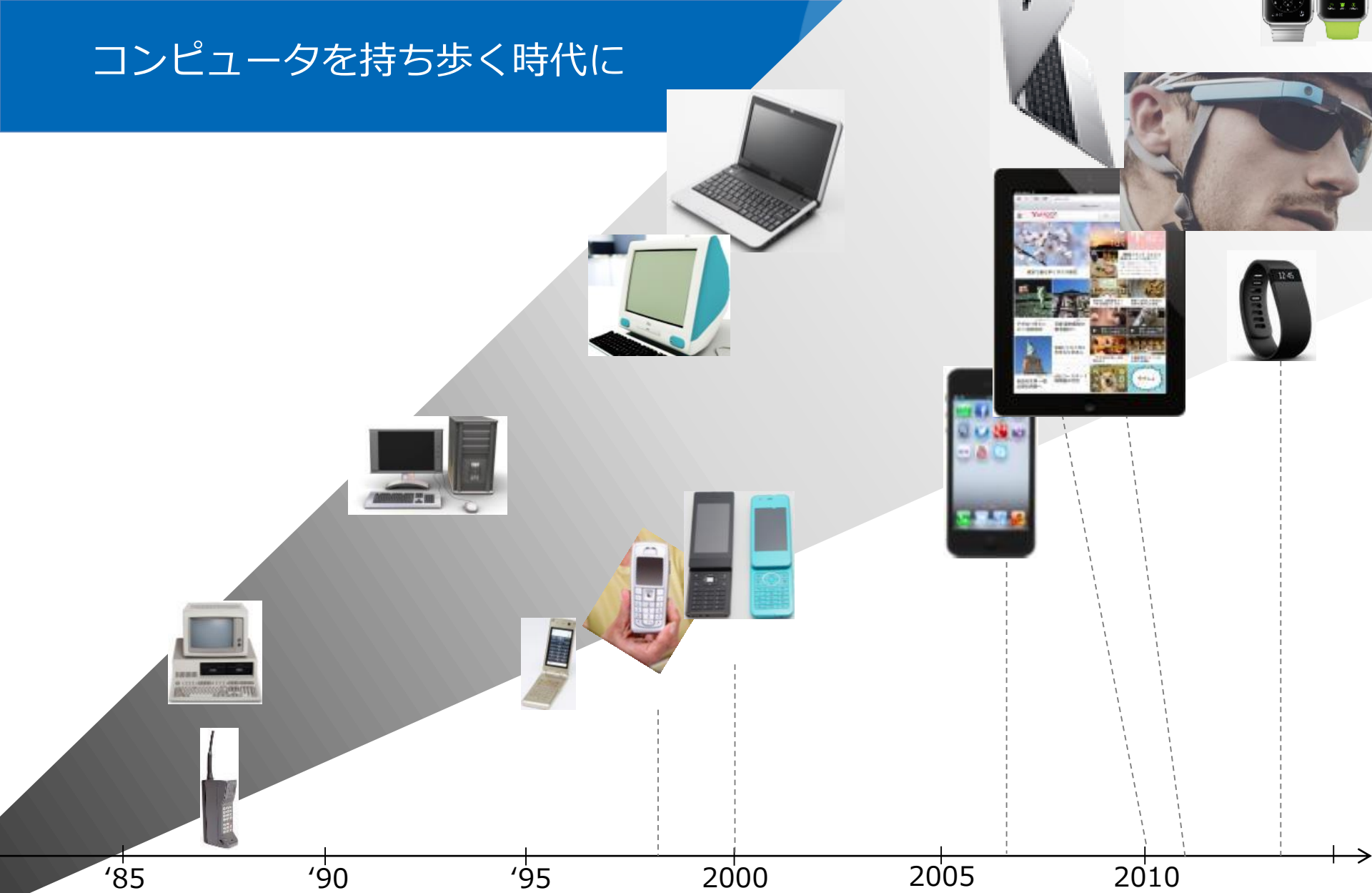
## 設立の目的

- この新しいデータプロフェSSIONナル（「データサイエンティスト」）に必要とされるスキルセットを定義する
- スキル育成と評価のための軸・基準を作る
- このデータプロフェSSIONナルの皆さんが相互に接し、やりとりできるオープンな環境を提供する
- 社会との対話を行なう



# データサイエンティスト とは何か？

# コンピュータを持ち歩く時代に

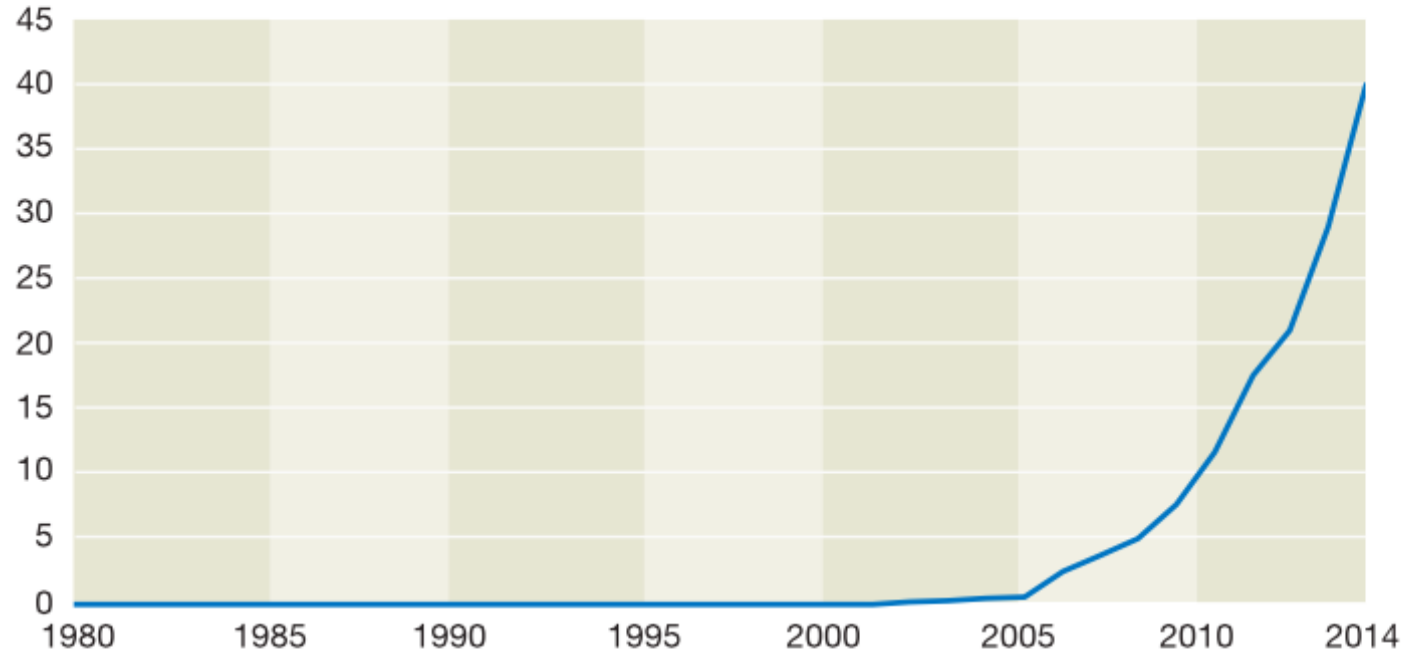


資料：安宅和人「データ時代に向けたビジネス課題とアカデミアに向けた期待」応用統計学セミナー2015.5.23  
(<http://www.applstat.gr.jp/seminar/ataka.pdf>)

モノからも含め  
多種多様なログデータが  
今後、爆発的に発生  
… ビッグデータ

# 計算キャパシティの爆増

**Annual additions to global business and consumer computing power, exaflops<sup>1</sup>**

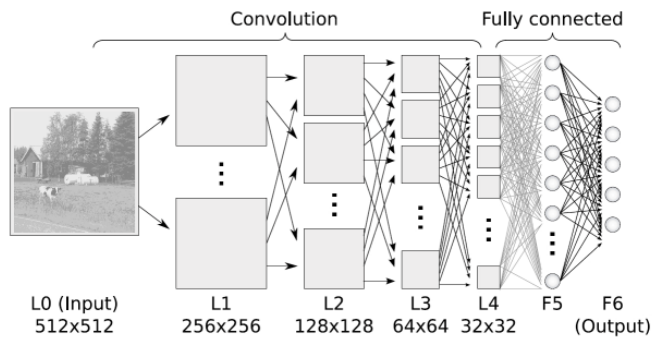


<sup>1</sup>An exaflop is 1 quintillion (10 to the 18th power) floating-point operations per second.

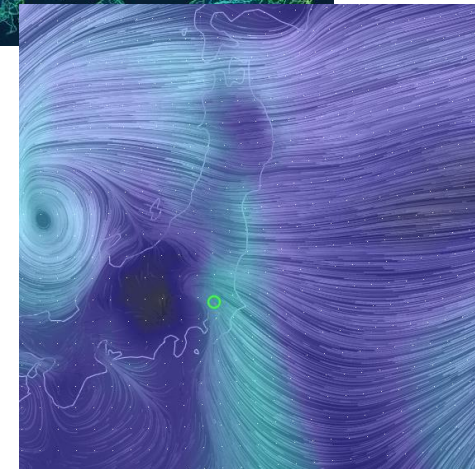
Source: William D. Nordhaus, "Two centuries of progress in computing," *Journal of Economic History*, 2007, Volume 67, Number 1, pp. 128–59; IDC; US Bureau of Economic Analysis; McKinsey analysis

資料 : McKinsey Quarterly 2014 Number3 p.16

## 大量データ処理技術 深層学習の実用化



## 可視化 (Data visualization) 技術の劇的な進展

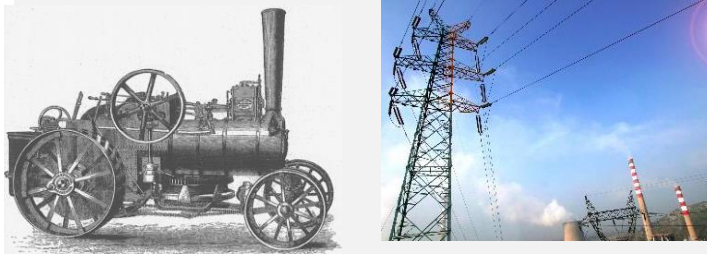


資料：東京大学 喜連川・豊田研HP、Universität Bonn, Institute for Computer Science HP、ヤフービッグデータレポートHP、Cameron Beccario「earth」WEBサイト

新しい  
リソース

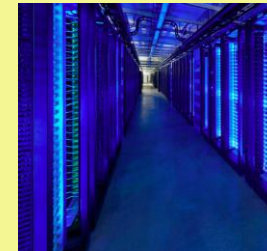
## 産業革命 (18~20世紀)

- 石炭と石油
- 蒸気機関
- 電気工学



## 情報産業革命 (現在)

- ビッグデータ
- 高い計算能力
- 情報科学の進化



本質

- 人間と家畜を肉体労働、手作業から開放する

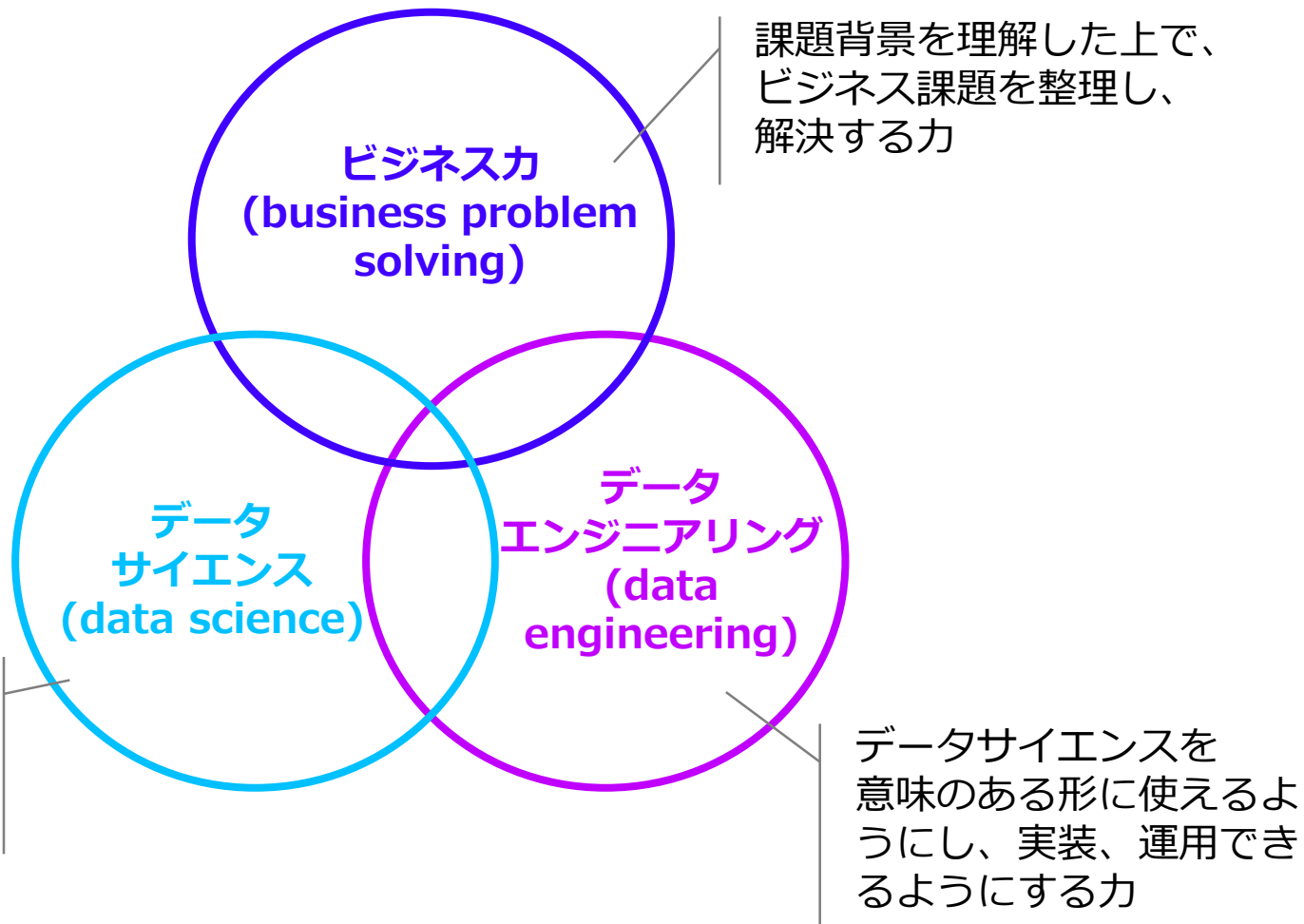


- 人間を退屈な数字入力、情報処理作業から開放する

資料：データサイエンティスト協会スキル委員会ディスカッション:安宅和人「データ時代に向けたビジネス課題とアカデミアに向けた期待」  
応用統計学セミナー2015.5.23 (<http://www.applstat.gr.jp/seminar/ataka.pdf>)

「データの持つ力を解き放つ」

# 3つのスキルセット



資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>



データサイエンス力、  
データエンジニアリング力をベースに  
データから価値を創出し、  
ビジネス課題に答えを出す  
プロフェッショナル

\* ここで「ビジネス」とは社会に役に立つ意味のある活動全般を指す

資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

# 単なるスペシャリストでもエキスパートでもない

## スペシャリスト

- 何らかの分野に特化した人

## エキスパート

- 何らかの分野について体系的で秀でた知識とスキルを持っている人

## プロフェッショナル

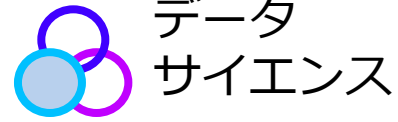
- 体系的にトレーニングされた専門性を持つスキルを持ち、
- それをベースに顧客（お客様、クライアント）にコミットした価値を提供し、
- その結果に対し、認識された価値の対価として報酬を得る人

資料：データサイエンティスト協会スキル委員会ディスカッション

# 課題解決のフェーズによって核となるスキルが変わる



ビジネス



データ  
サイエンス



データ  
エンジニアリング

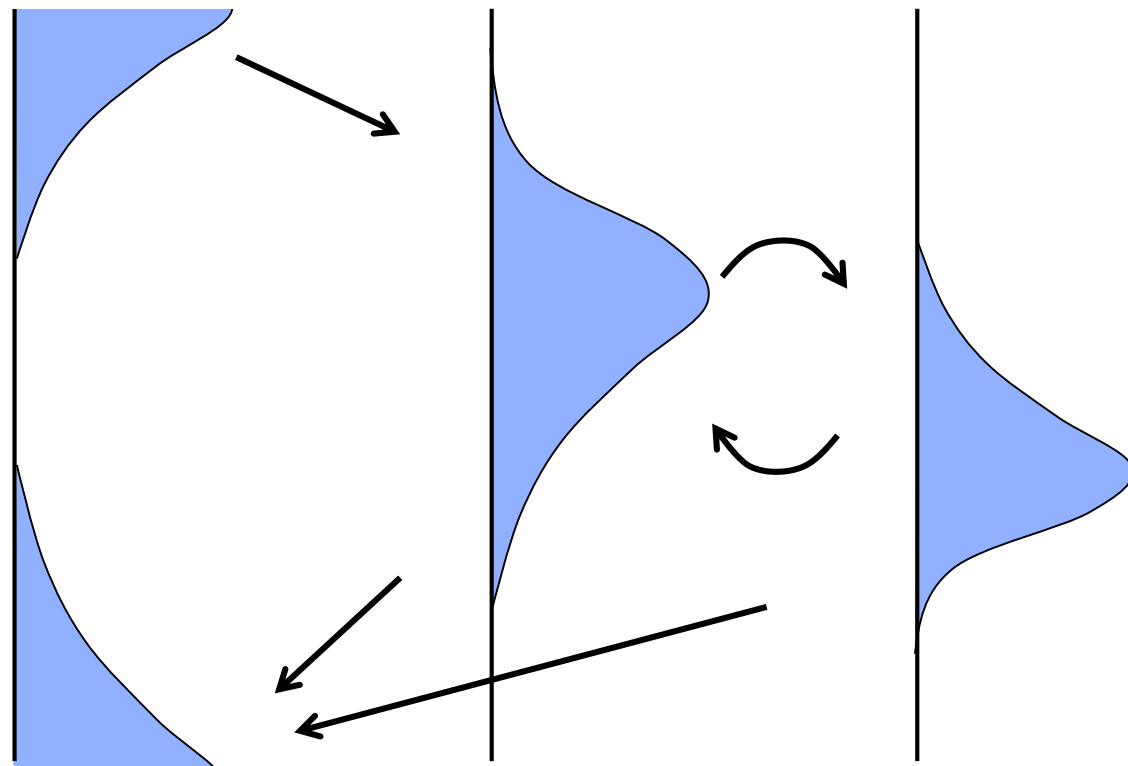
目的・テーマ設定

問題定義

アプローチの設計

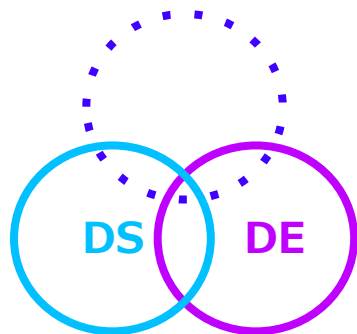
処理・分析

解決



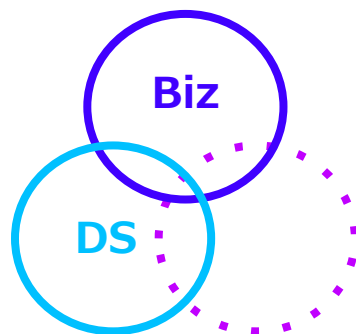
資料 : データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

# どの一つが欠けてもダメ、、、優位はない



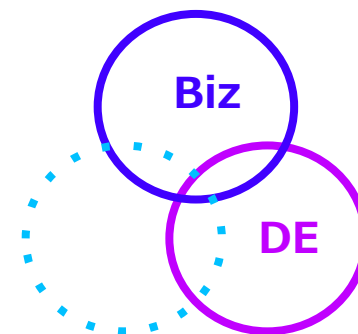
データスペシャリスト  
ではあるが、そもそも  
解決すべき問題が定義、  
整理できない

> プロフェッショナル  
ではない



ビジネス課題もわかり  
そこでのサイエンスの  
利活用もわかるが実装  
できない

> 必要な変化を起こせ  
ない



ビジネス課題の上で、  
実装を用意できるが、  
かなめとなるサイエン  
スの知恵が足りない

> 賢いやり方を提供で  
きない

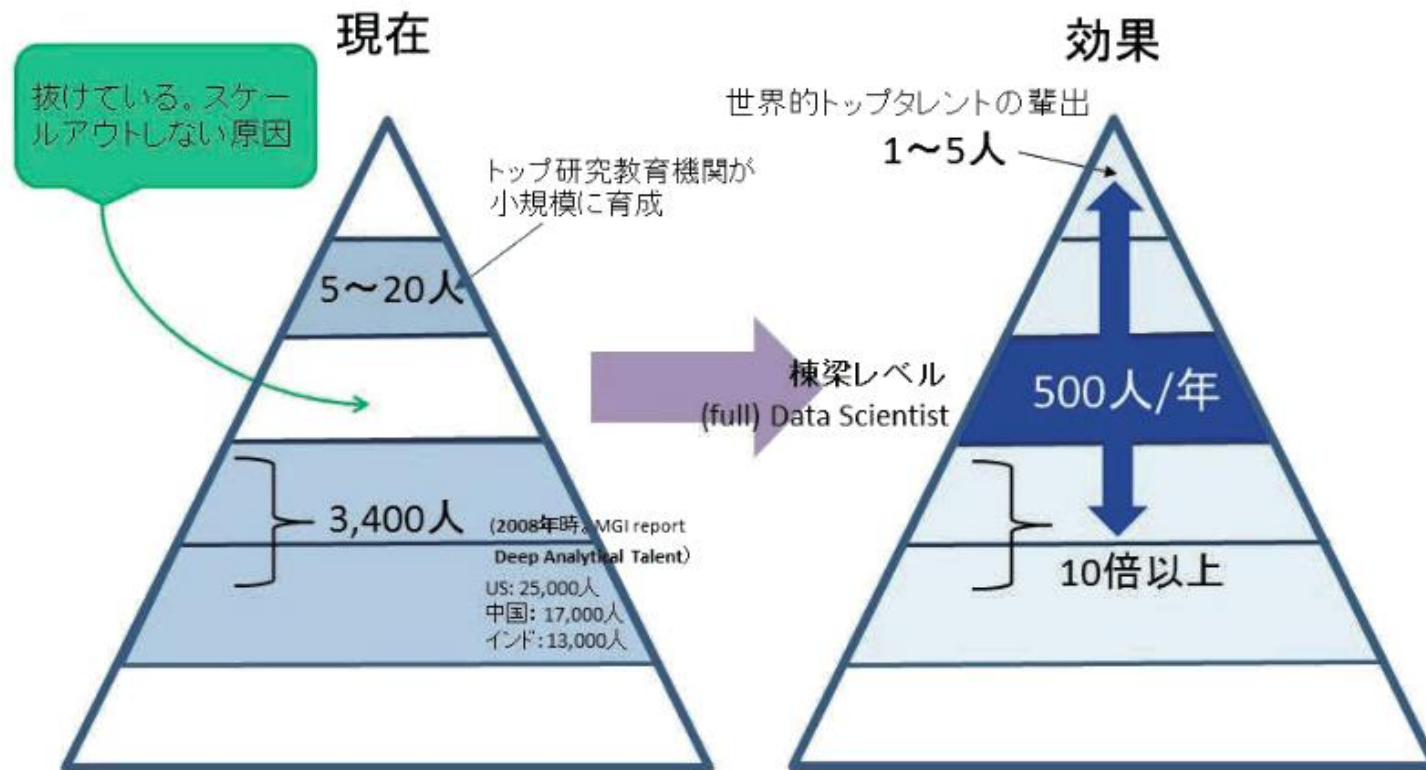
# スキルのレベル感は？

# スキルレベル

	スキルレベル	目安	対応できる課題
データサイエンティスト	シニア・データサイエンティスト	<ul style="list-style-type: none"><li>業界を代表するレベル</li></ul>	<ul style="list-style-type: none"><li>産業領域全体</li><li>複合的な事業全体</li></ul>
	フル・データサイエンティスト	<ul style="list-style-type: none"><li>棟梁レベル</li></ul>	<ul style="list-style-type: none"><li>対象組織全体</li></ul>
	アソシエート・データサイエンティスト	<ul style="list-style-type: none"><li>独り立ちレベル</li></ul>	<ul style="list-style-type: none"><li>担当プロジェクト全体</li><li>担当サービス全体</li></ul>
	アシスタント・データサイエンティスト	<ul style="list-style-type: none"><li>見習いレベル</li></ul>	<ul style="list-style-type: none"><li>プロジェクトの担当テーマ</li></ul>
一般人	データ使い	<ul style="list-style-type: none"><li>賢くデータを器用にする人</li></ul>	<ul style="list-style-type: none"><li>担当業務</li></ul>
	普通の人	<ul style="list-style-type: none"><li>特になし</li></ul>	

資料：データサイエンティスト協会プレスリリース (2014.12.10) <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>

# とりわけ棟梁レベルが足りていない



資料：情報・システム研究機構 『ビッグデータの利活用のための専門人材育成について』（ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会）2015.7.15

## 1. 昨年のrecap

- ① 本協会設立の背景と目的
- ② データサイエンティストとは何か？
- ③ 3つのスキルとスキルレベル

## 2. スキルチェックリスト

- ① 全体として（昨年 vs 今年）
- ② 個別領域の広がり
- ③ レベル感の考え方

## 3. 実行に向けて

- ① チームをどう作るか？
- ② 人をどう育てるか？
- ③ 目指す人はどう着手すべきか？



## 気づき

## ポイント

データ  
サイエンス  
力

- 情報科学も情報系、機械学習系、データ可視化系いずれかだけでは足りない

- 総合的な解決アプローチにつなげる力がより必要。分野別の専門性はnice to haveにすぎない

データ  
エンジニア  
リング力

- いわゆるプログラマー、SIer的なエンジニアの通常技術だけでは足りない

- 既存のエンジニアリング業務を超え、ビッグデータ処理を実現できる必要がある

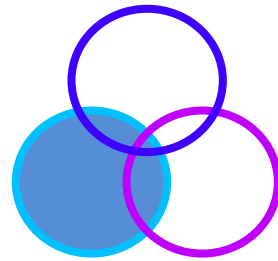
ビジネス力

- 問題解決力は単なる領域での知見やコンサル的なものを超える必要がある

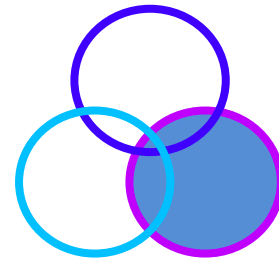
- 他のスキル領域と話せる力を含めたデータ視点での課題解決力を持つ必要がある

# スキル表

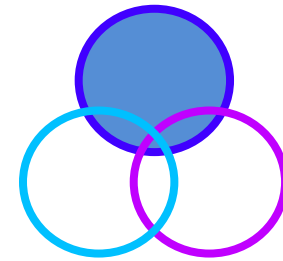
(項目数)



データ  
サイエンス



データ  
エンジニアリング



ビジネス

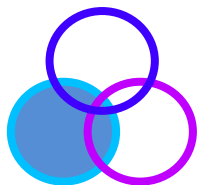
レベル  
計

★★★★	棟梁レベル (full)	58	37	44	(139)
★★★	一人前レベル (Associate)	68	43	49	(160)
★	見習いレベル (Assistant)	54	39	30	(123)
	領域計	180	119	123	(422)

- データセキュリティはデータエンジニアリングとビジネスそれぞれの視点から扱っています
- 以下の項目は本チェックリストではカバーしていません
  - ✓ 特定のスキル項目ができると当然できると考えられるスキル
  - ✓ 特定の業界で重要であっても、一般性が低いスキル
  - ✓ 特定作業ができるために必要になるSAS、SPSSなど具体的ツールを使うスキル\*
- 20名弱の実務家の協議による検討結果 (ver.1) であり、必要に応じ、適宜、見直していく位置づけです
- 個別の項目ができるできないの判断は、いまのところ各自のご判断で行っていただく前提です (後述)

\* ただし具体名を入れないと理解が困難なケースでは固有名詞を入れてあります

# 領域の広がり（データサイエンス）



解析技術

探索的

サンプリング

データ加工

検定/判断

Static

機械学習

グルーピング

性質・関係性の把握

パターン発見

グラフィカルモデル

Dynamic

予測

時系列分析

シミュレーション  
/データ同化

Data Visualization

非構造化  
データ処理

言語処理

画像処理

音声処理

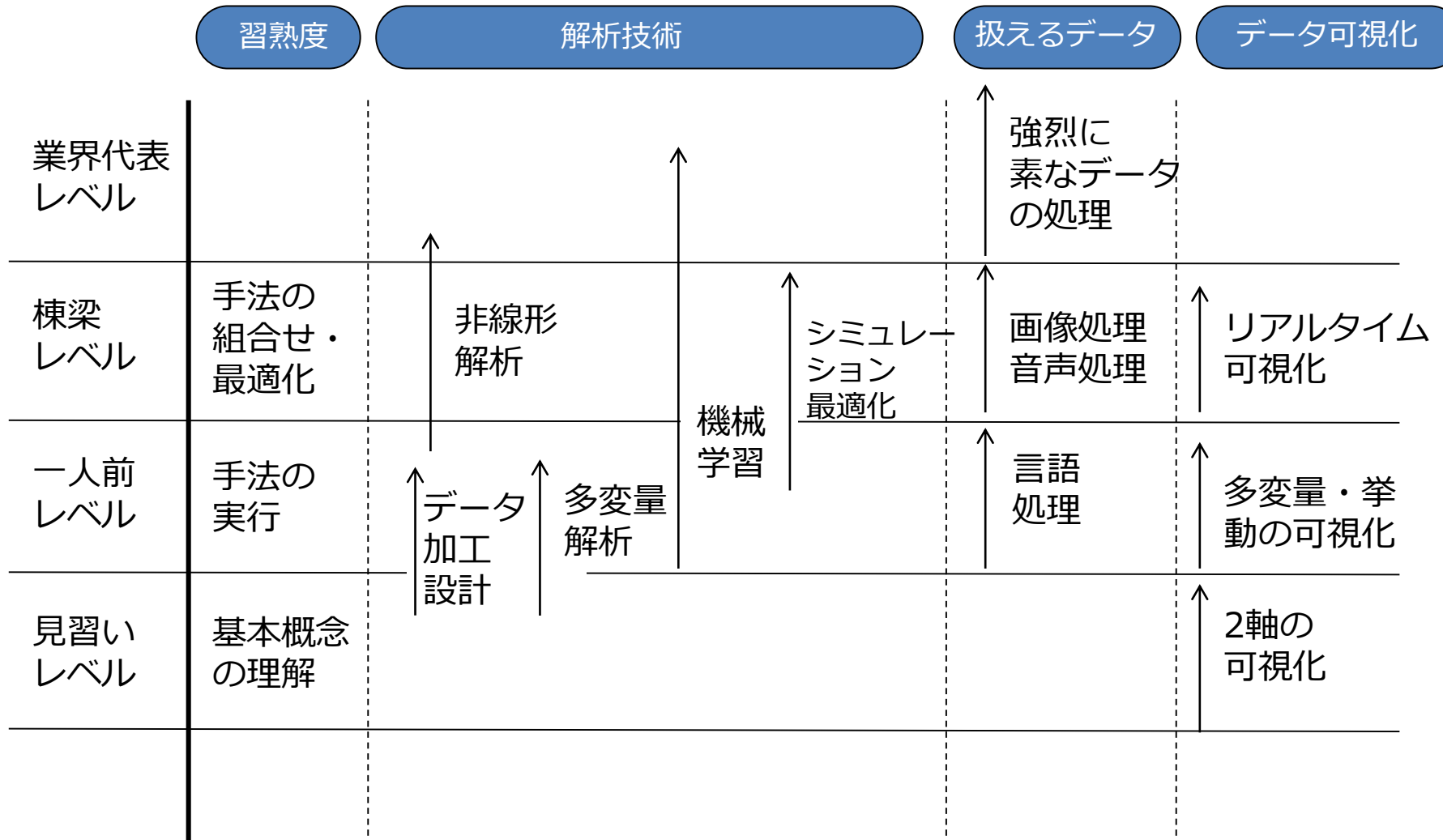
基礎技術

統計数理基礎

統計数理応用

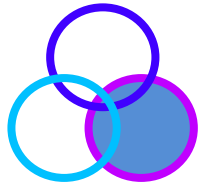
# レベル感の目安 (データサイエンス)

2015.11.13 現在



資料：データサイエンティスト協会スキル委員会討議

# 領域の広がり (データエンジニアリング)



守る技術

ITセキュリティ

実装技術

データ  
収集

データ  
構造

データ  
蓄積

データ  
加工

データ  
共有

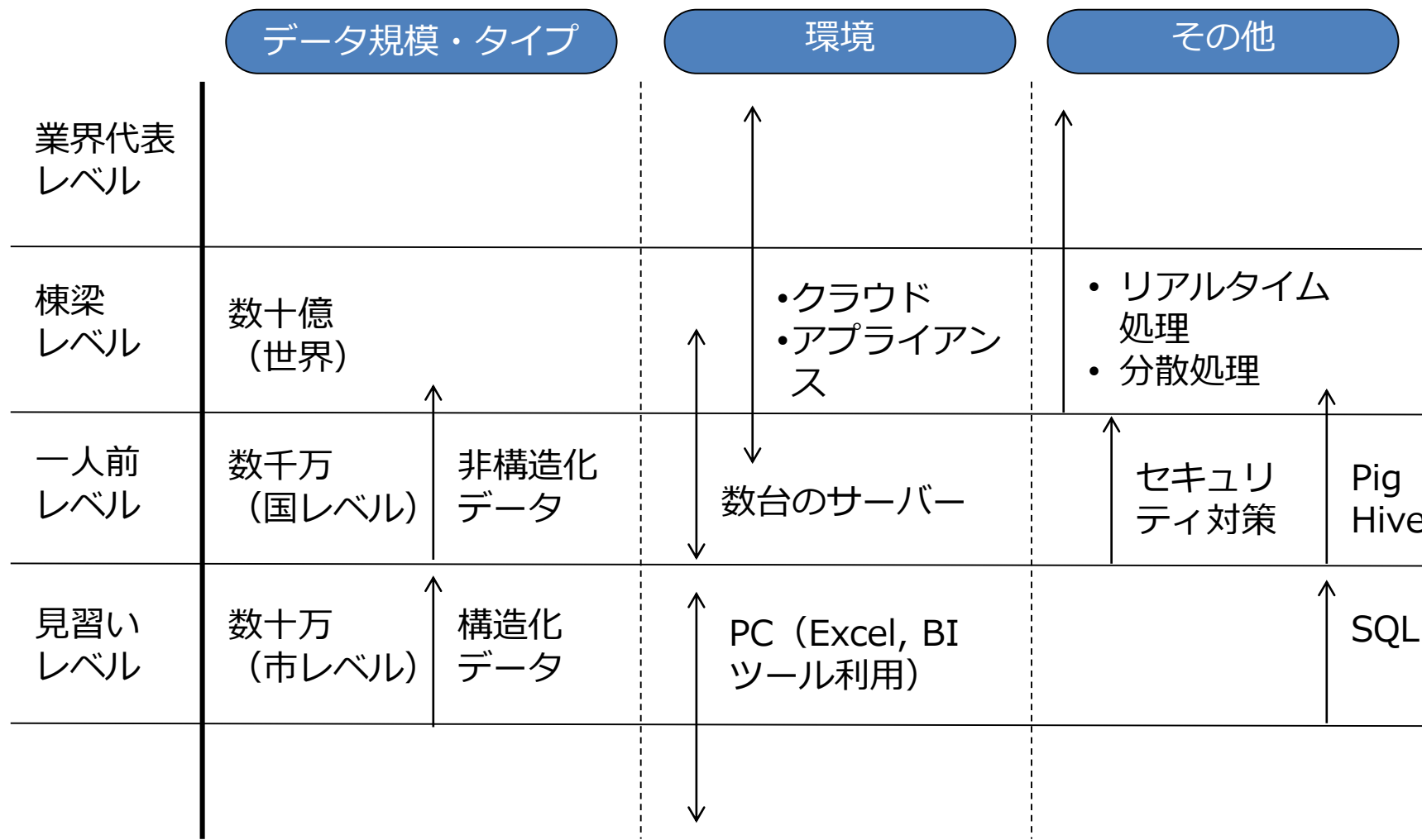
環境構築

基礎技術

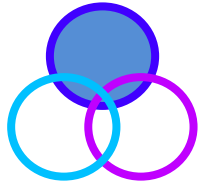
プログラミング

# レベル感の目安 (データエンジニアリング)

2015.11.13 現在



資料：データサイエンティスト協会スキル委員会討議



ビジネス  
課題解決

活動マネジメント

事業に実装する

データ  
課題解決

データ  
入手

データの  
理解・検証

意味合いの  
抽出・洞察

プロセス

基礎能力

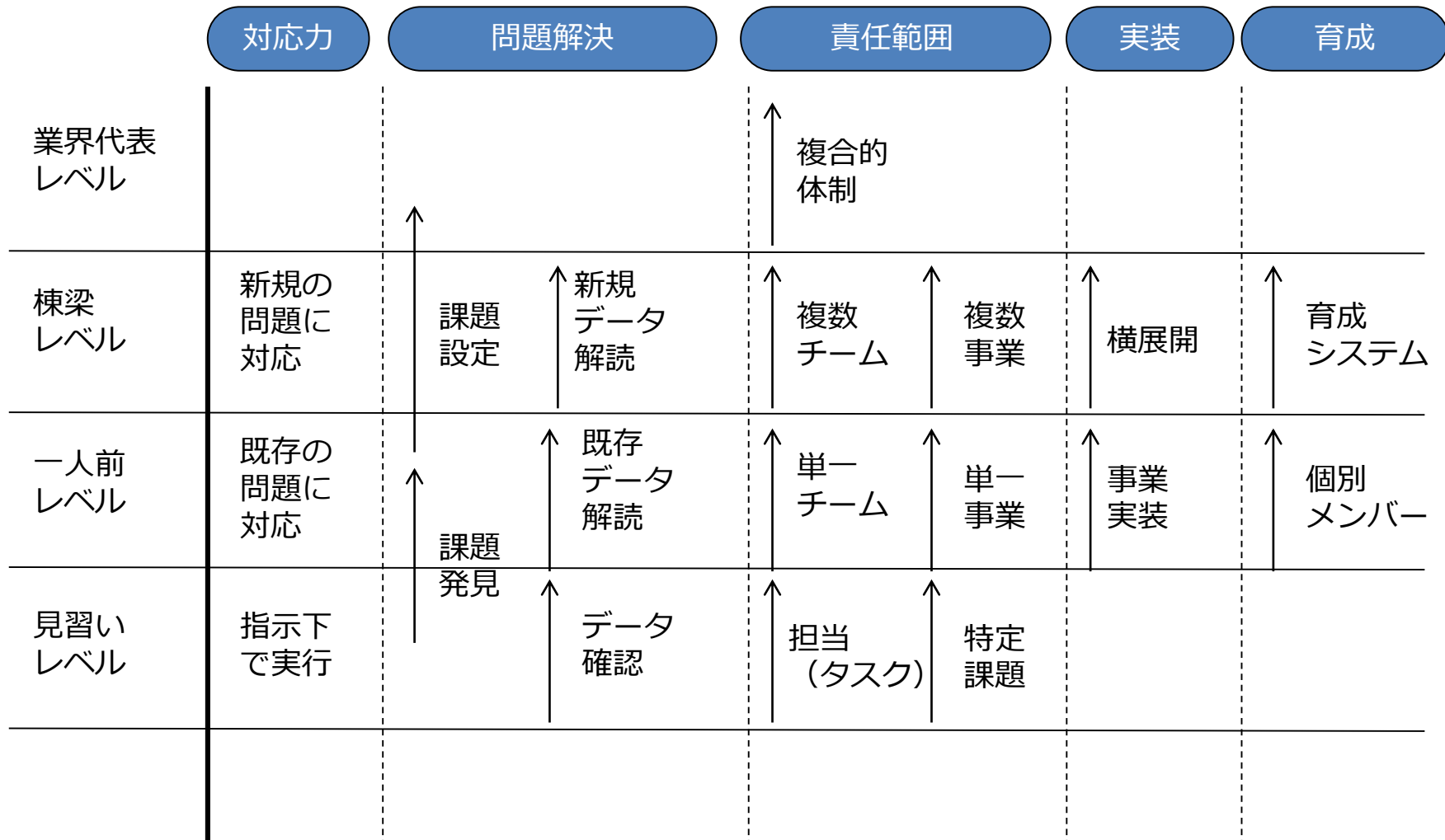
論理的思考

行動規範

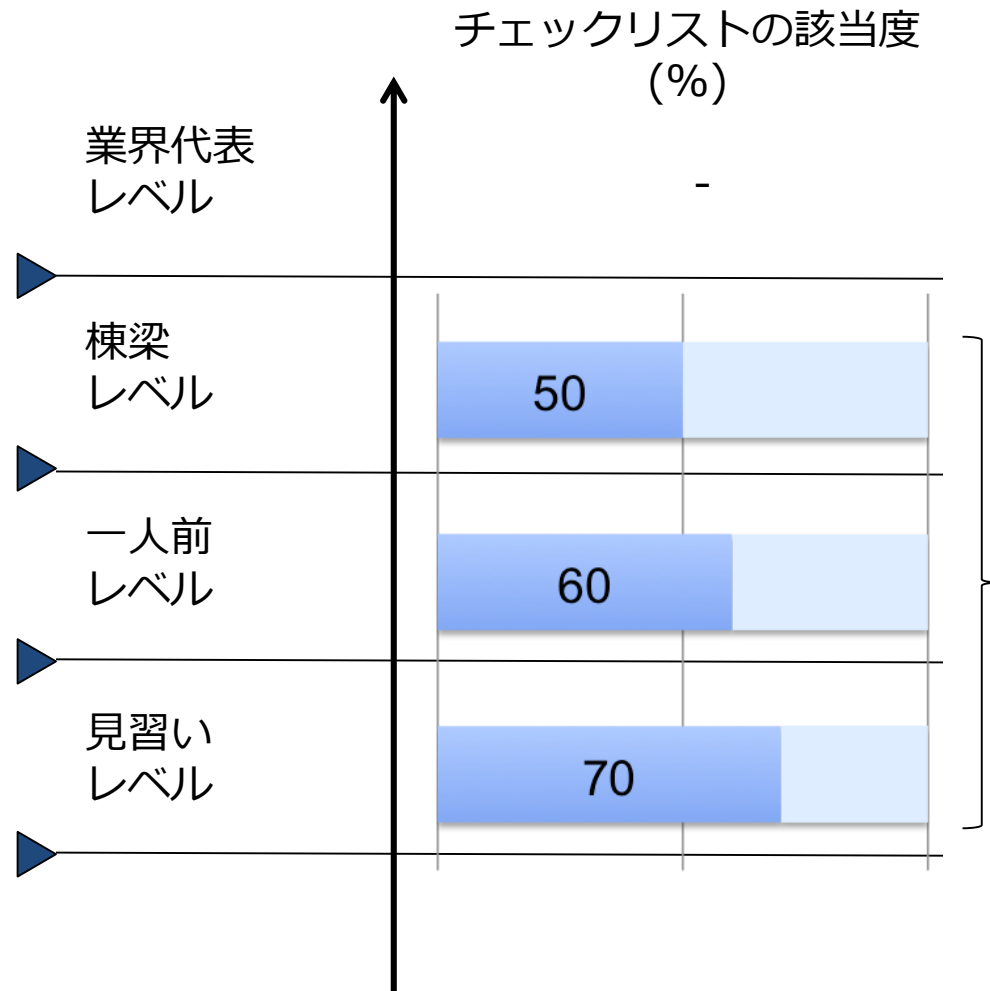


# レベル感の目安（ビジネス）

2015.11.13 現在



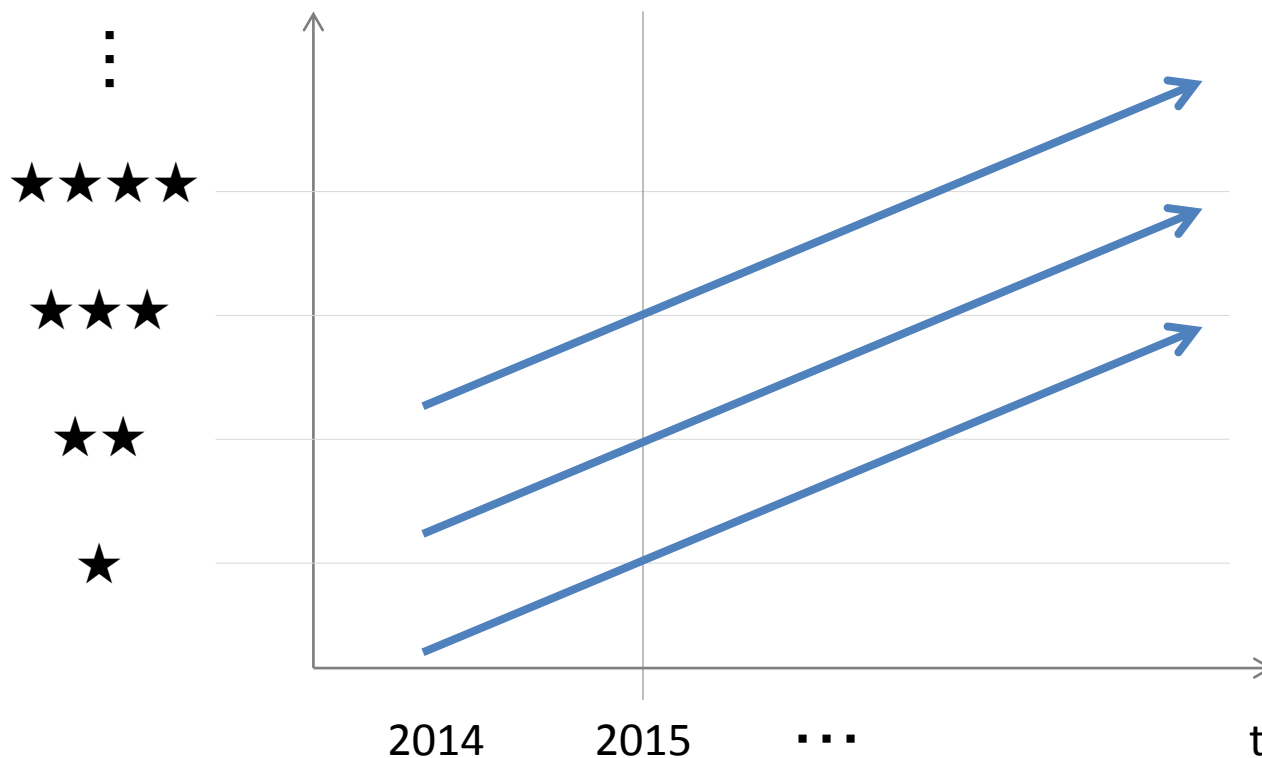
資料：データサイエンティスト協会スキル委員会討議



- 3つの領域別に判定
- 各レベルの必須項目を一つでも満たさないとアウト
- 下のレベルは満たしていることが前提

# スキルは本質的にmoving target

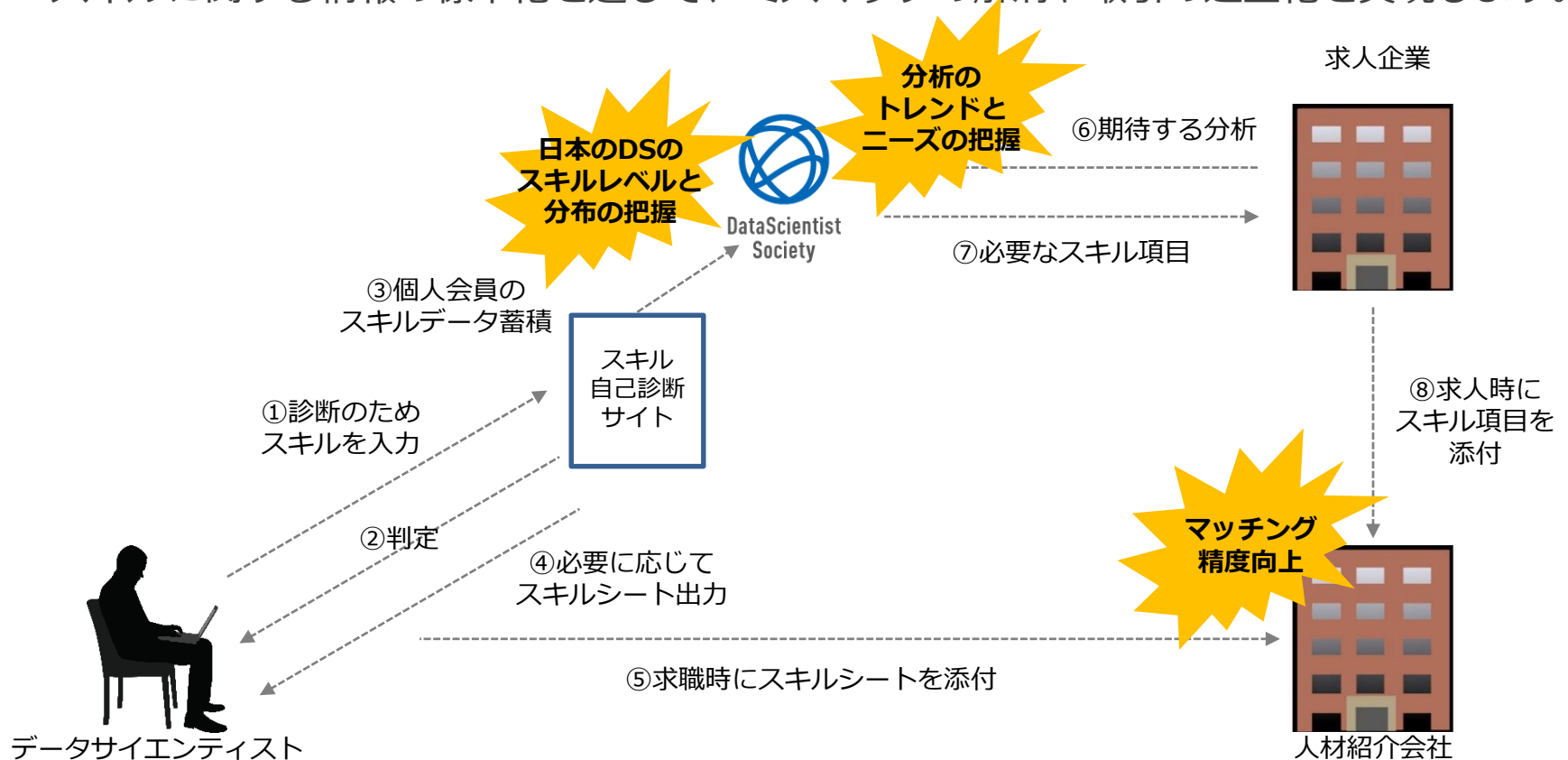
CONCEPTUAL



資料：データサイエンティスト協会スキル委員会討議

# スキルチェックの仕組み（イメージ）

- 2016年度にデータサイエンティスト自身がスキルのチェックをサイト上で実施できる自己診断の仕組みを公開予定。  
人材紹介業界でのスタンダードを目指し、データサイエンティストや企業間でのスキルに関する情報の標準化を通じて、ミスマッチの解消や取引の適正化を実現します。



## 1. 昨年のrecap

- ① 本協会設立の背景と目的
- ② データサイエンティストとは何か？
- ③ 3つのスキルとスキルレベル

## 2. スキルチェックリスト

- ① 全体として（昨年 vs 今年）
- ② 個別領域の広がり
- ③ レベル感の考え方

## 3. 実行に向けて

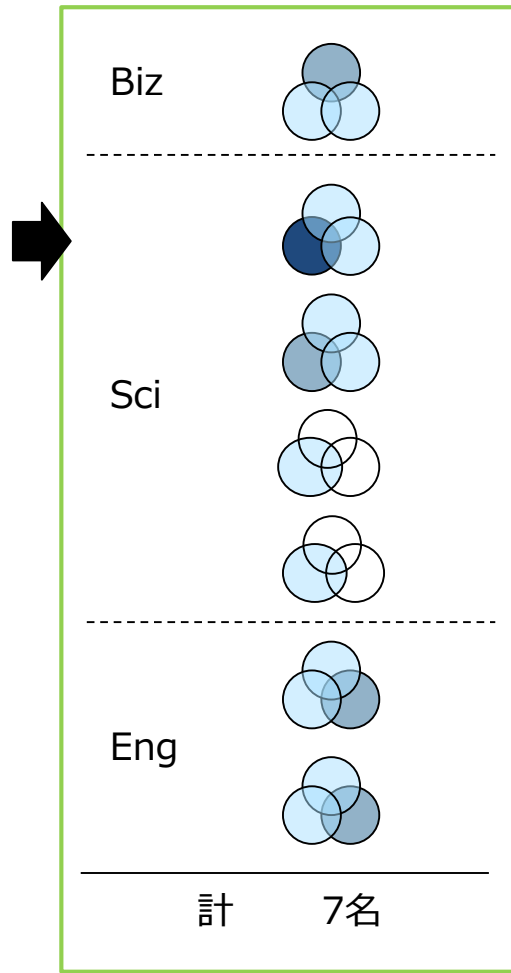
- ① チームをどう作るか？
- ② 人をどう育てるか？
- ③ 目指す人はどう着手すべきか？

- 具体的にどう回すか？
- 具体的にどうやって育成していくべきか？
- 目指す人は何をどうしていったらいいのか？

実際のチームは  
どう組んだらいいのか？

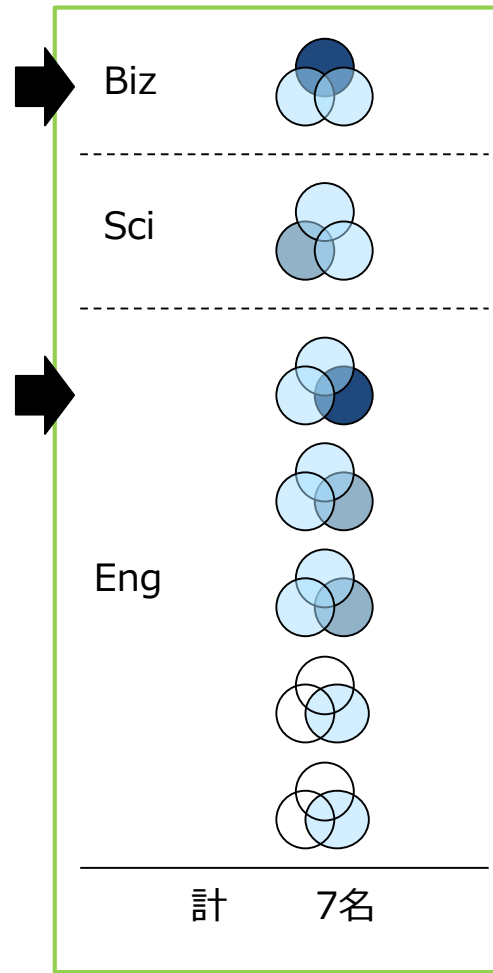
# チーム編成事例 (1/3) : ウェブサービス提供

例1 : レコメンド (Sci寄り)

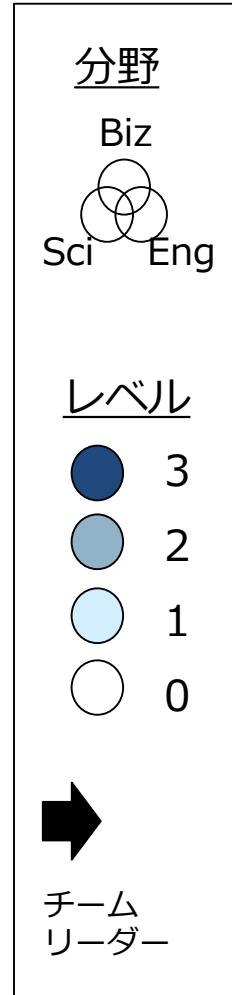


3ヶ月(改善)~12ヶ月(新規)

例2 : アクセス解析 (Eng寄り)



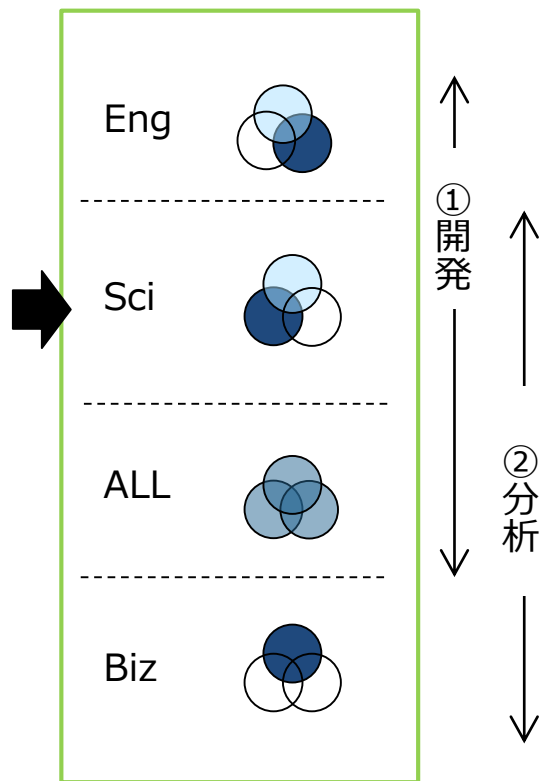
1ヶ月(改善)~6ヶ月(新規)



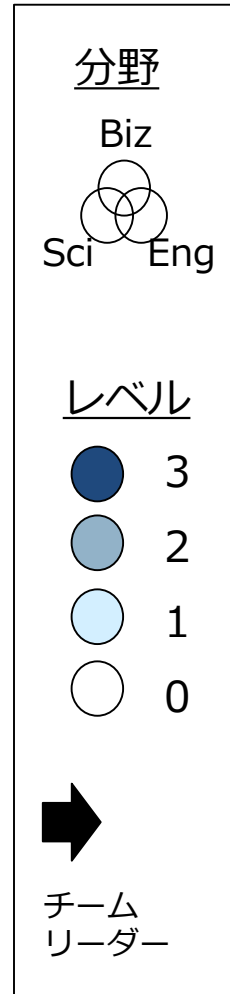
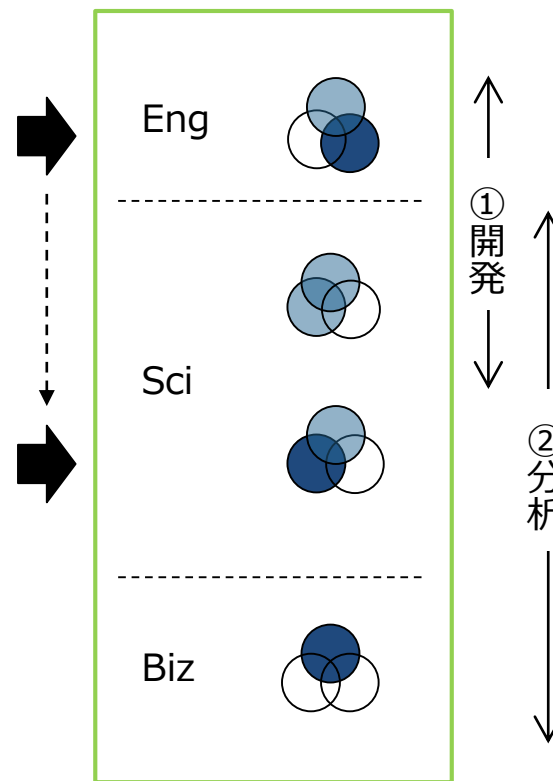


# チーム編成事例(2/3) : マーケティングソリューション

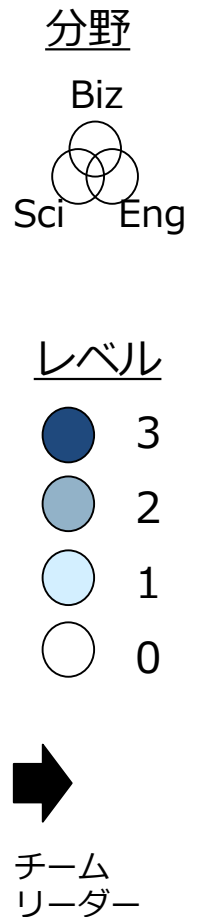
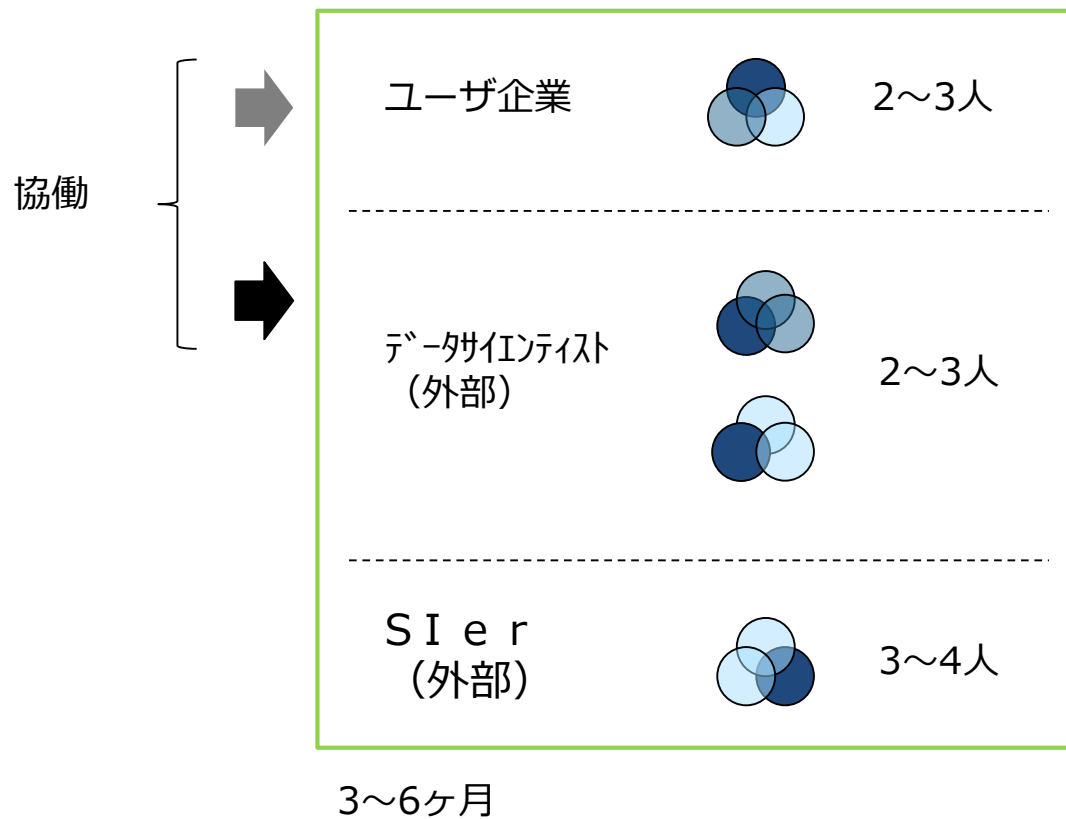
## 例3 : 行動データ解析



## 例4 : マーケティング投資最適化



## 例5：マーケティングオートメーション導入

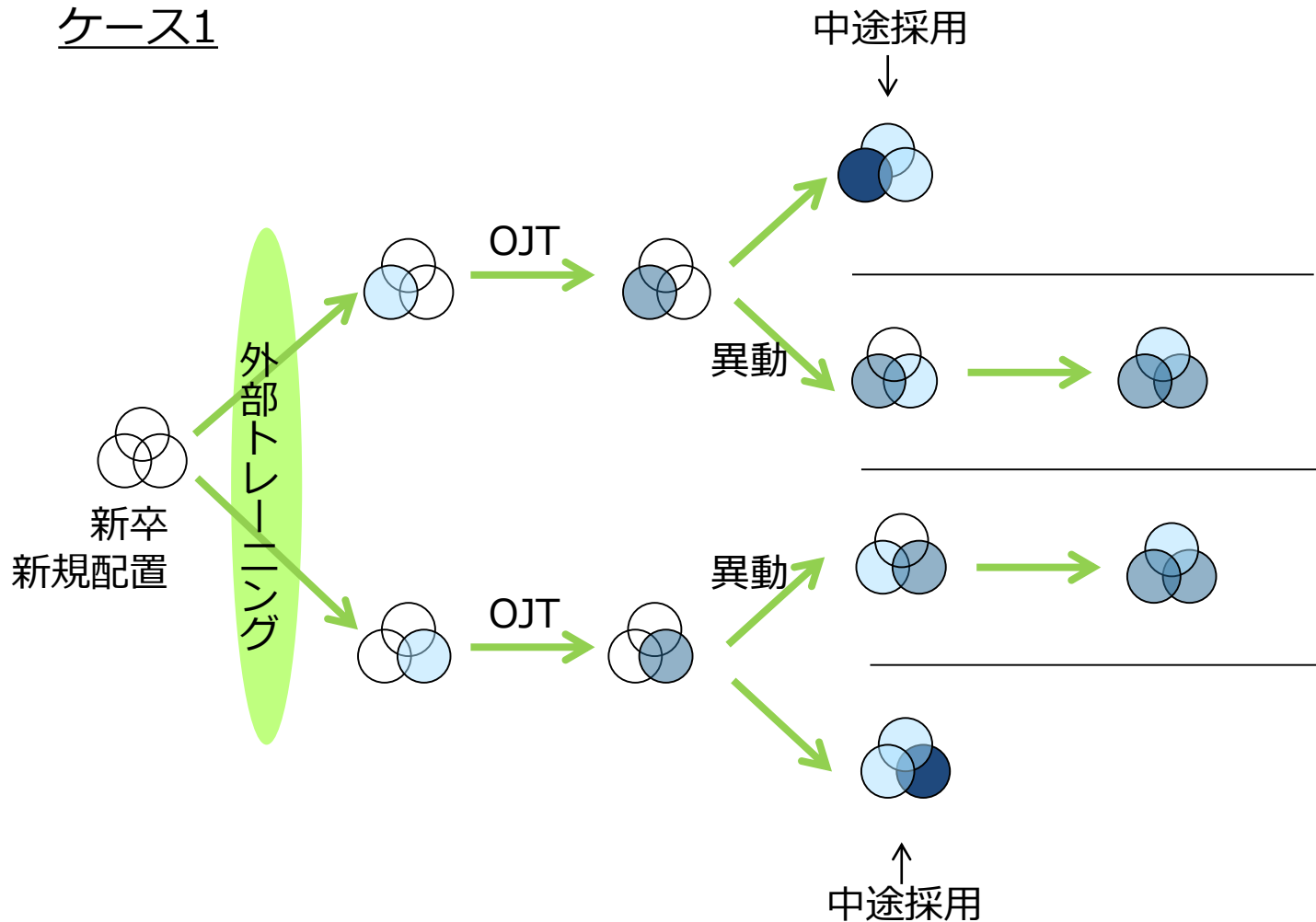


- 単独の一人ではなく、チームとして必要なスキルセットを実現するのが現実解
- 課題解決の重さがどこにあるかで組成も変化（サイエンスより、エンジニアよりなど）
- 開発、運用などのフェーズによっても異なる
- チームリーダーはコアになる課題解決と俯瞰できる力の有無でさまる

- そもそもこのような体制を作れない会社はどうしたらいいのか？
- どういうヒトの組み合わせ、体制にしたらいいのかわからない会社はどうしたらいいのか？

ではスキル育成は  
どうしたらいいのか？

## ケース1



**分野**

Biz

Sci Eng

**レベル**

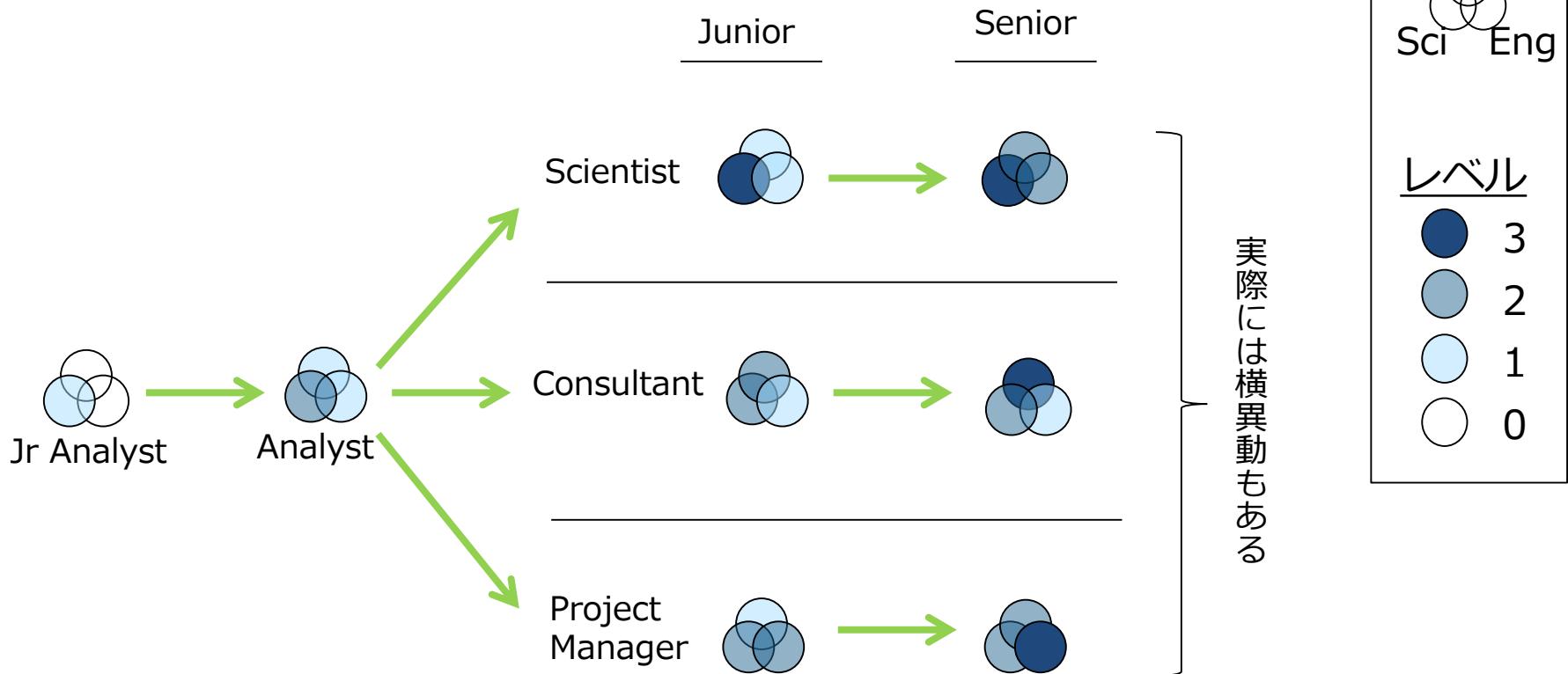
● 3

● 2

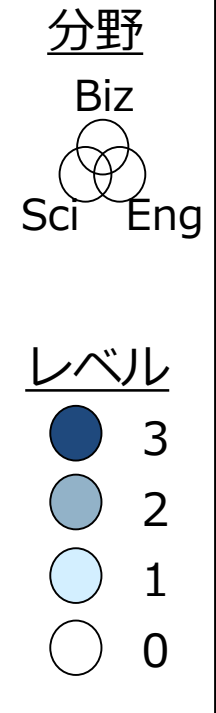
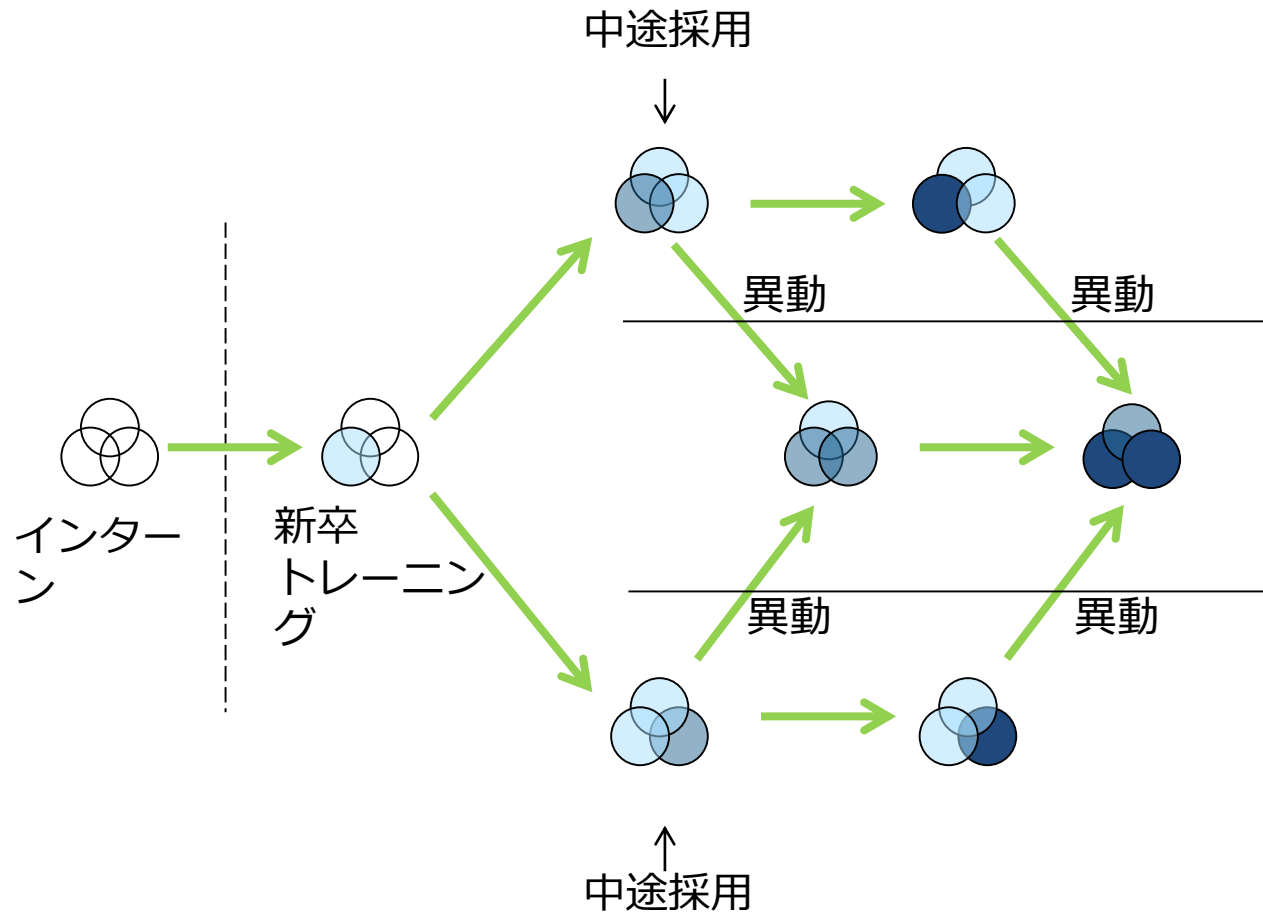
● 1

○ 0

## ケース2

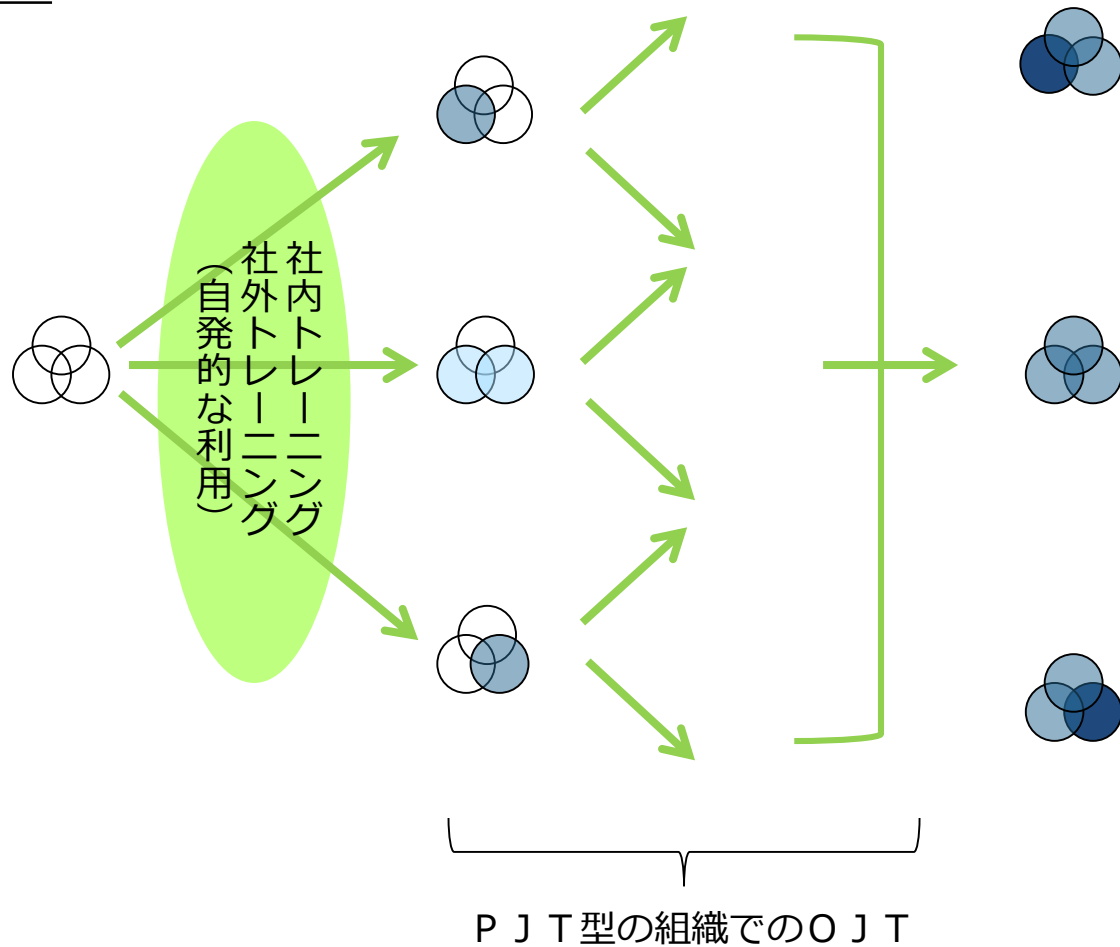


## ケース3





## ケース4



- 見習いレベルまでは共通的なトレーニングで、そこから先は実践を通じてが基本
- サイエンス、エンジニア両方のスキルを同時というより、プロジェクトや部門異動によって育成していくのが現実解
- ビジネス力は複数の領域を超えて仕事を始め、全体を俯瞰してマネージする中で育成するケースが多い

- 特定の領域の職人的な環境にいるひと（特にエンジニア上がりの人）はどうやってデータサイエンティストになるのか？
- より高度化していくスキルにどのように対応していくべきなのか？
- 異動を自分でコントロールしようがない時にどのように新しいスキルを身につけていくべきなのか？
- 高等教育機関（特に大学）はどう活用したら良いのか？

「データの持つ力を解き放つ」