

2016年1月21日  
日本アイ・ビー・エム株式会社

一般社団法人オープン&ビッグデータ活用・地方創生推進機構 御中  
公共施設集約化推進施策の一環としての  
コミュニティバス最適配置の可能性検証プロジェクト

中間報告資料



1. 本プロジェクトにおける前提
2. プロジェクトの概要と目的
  1. プロジェクト概要
  2. 目的とゴール
3. プロジェクトのアプローチ
  1. アプローチ
  2. スケジュール
  3. 体制・役割分担
4. 公共施設高度化検討におけるデータ活用検討のポイント
  1. 全体像
  2. データ活用検討のポイント
    1. 分析視点・対象施設の検討
    2. 利用者アンケート調査・設計
    3. 調査実施・データ化作業
    4. SPSSを用いた統計解析・テキストマイニング
    5. データアナリストへのトレーニング
    6. 集約対象候補選定・コストシミュレーション
5. 自治体におけるデータ活用促進に向けた考慮点・気づき
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題
6. 参考資料
  1. 別紙1：公共施設アンケート調査票
  2. 別紙2：川越市公民館利用者アンケート調査結果
  3. 別紙3：SPSSトレーニングセッション資料
  4. 別紙4：コストシミュレーションツールの内容と分析結果

## 本プロジェクトにおける前提

---

- 本プロジェクトにおける資料の記述や検討内容に先立って、以下の2点を前提としてご確認ください。
  - **本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。**
  - 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために**検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。**

- 本プロジェクトは、**新たな価値・サービス・ビジネス創出における公共データ利活用の有効性および、他の自治体への展開可能なリファレンス事例・示唆を得る事を主眼**として実施しました。
- 具体的には、川越市役所 データアナリストのご協力を得て、公共施設利用者アンケートを中心としたデータ解析を通じて、施設利用状況・利用者属性・施設満足度との相関等を分析し、公共施設高度化に向けた検討を行いました。
- この中で行った、SPSSテキストマイニング機能を活用した利用者アンケート分析作業等を通じ、「**自治体におけるデータ活用やデータアナリスト育成における考慮点・課題**」を抽出を行ったものです。

## 川越市における公共施設の 利用状況を明らかにする

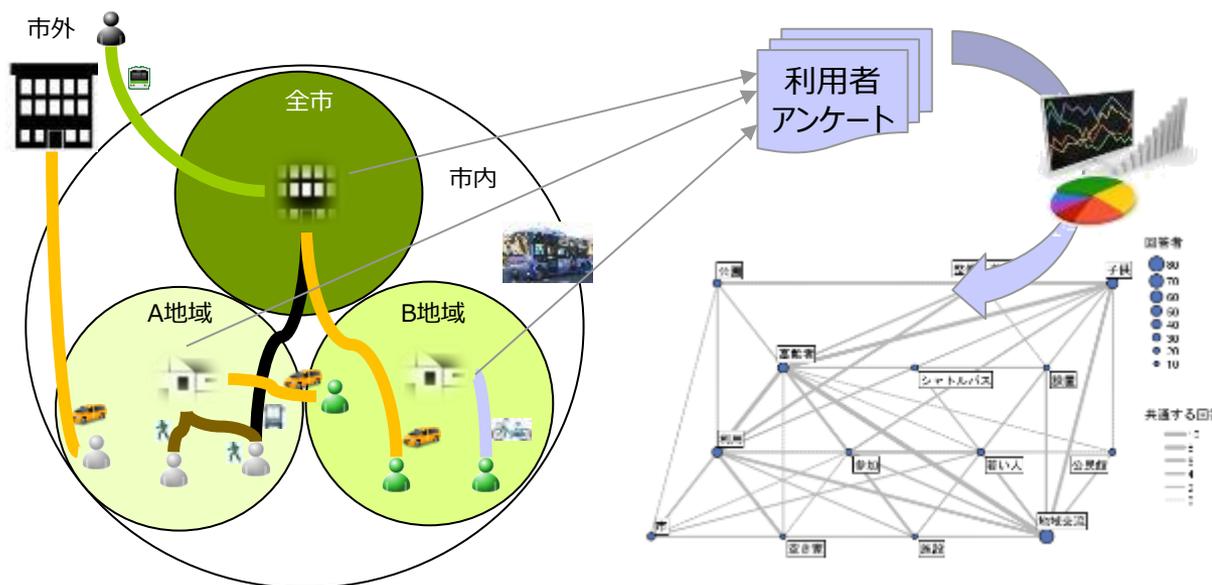
- 川越市提供のデータおよび利用者アンケート調査を通じて、公共施設利用の現状および利用者ニーズを分析

## 利用者の生の声を分析することで 高度化に向けた示唆を得る

- 利用者アンケートの自由記述欄をテキストマイニングにより解析、回答者属性と合わせて分析することで、高度化に向けた考察・示唆を導出

## 公共施設高度化、 および今後のデータ活用に 向けた示唆を得る

- 利用者アンケートの自由記述欄をテキストマイニングにより解析、回答者属性と合わせて分析することで、高度化に向けた考察・示唆を導出



1. 公共施設高度化による住民サービス向上への示唆
2. 公共施設に係る費用対効果の改善への示唆

## 自治体におけるデータ活用やデータアナリスト育成における考慮点や工夫・課題の抽出

3. 本取り組みの広域連携事業への将来的な展開に向けた示唆
4. 政策決定におけるデータ活用に向けたリファレンス事例獲得

## 目的

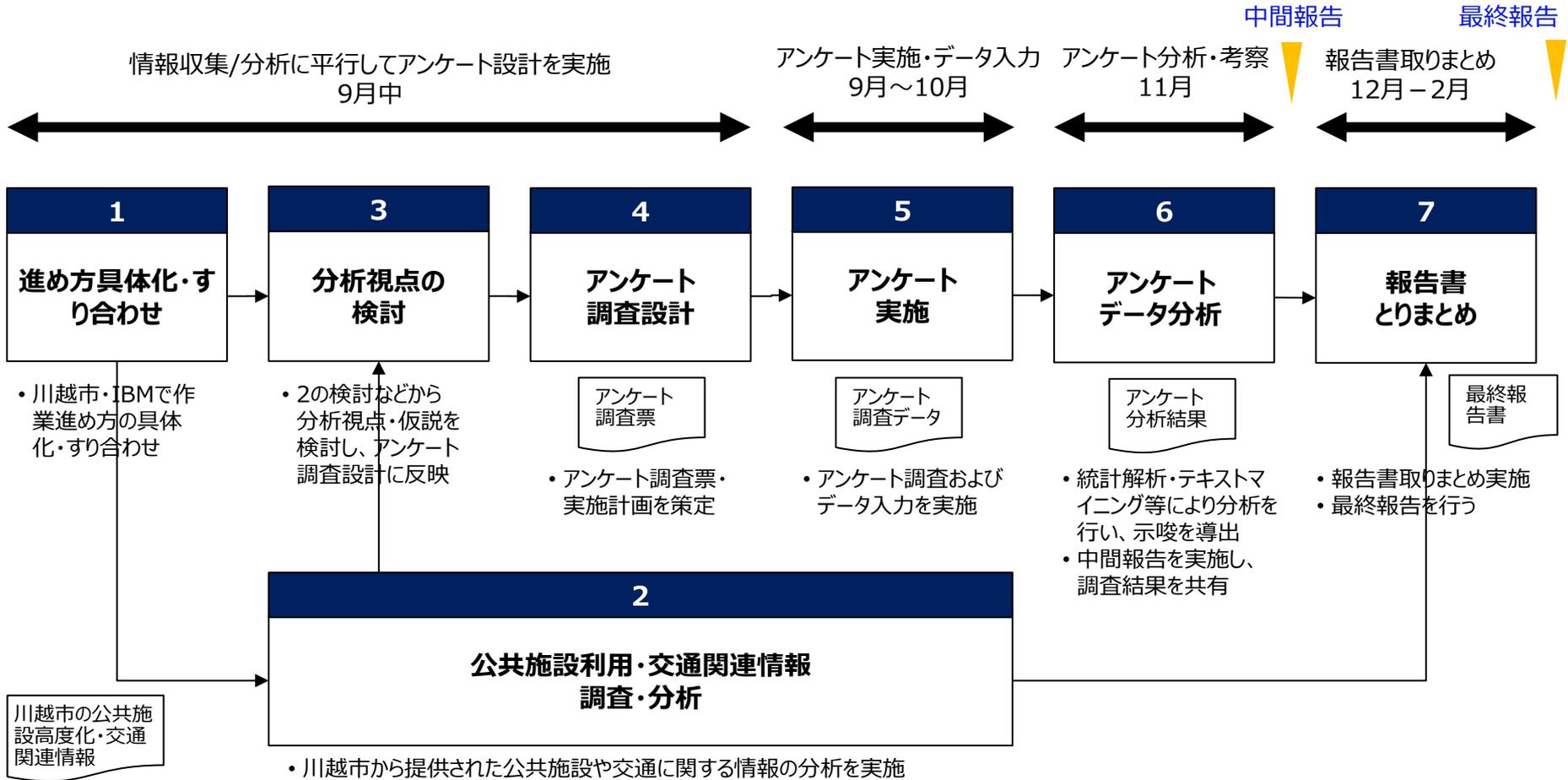
- 川越市の公共施設利用者アンケートを中心としたデータ解析を通じて、公共施設利用状況・利用者属性・施設満足度の相関等を分析し、公共施設高度化に向けた示唆を得る。
- この中で行った、SPSSテキストマイニング機能を活用した利用者アンケート分析作業等を通じ、「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出する。

## ゴール

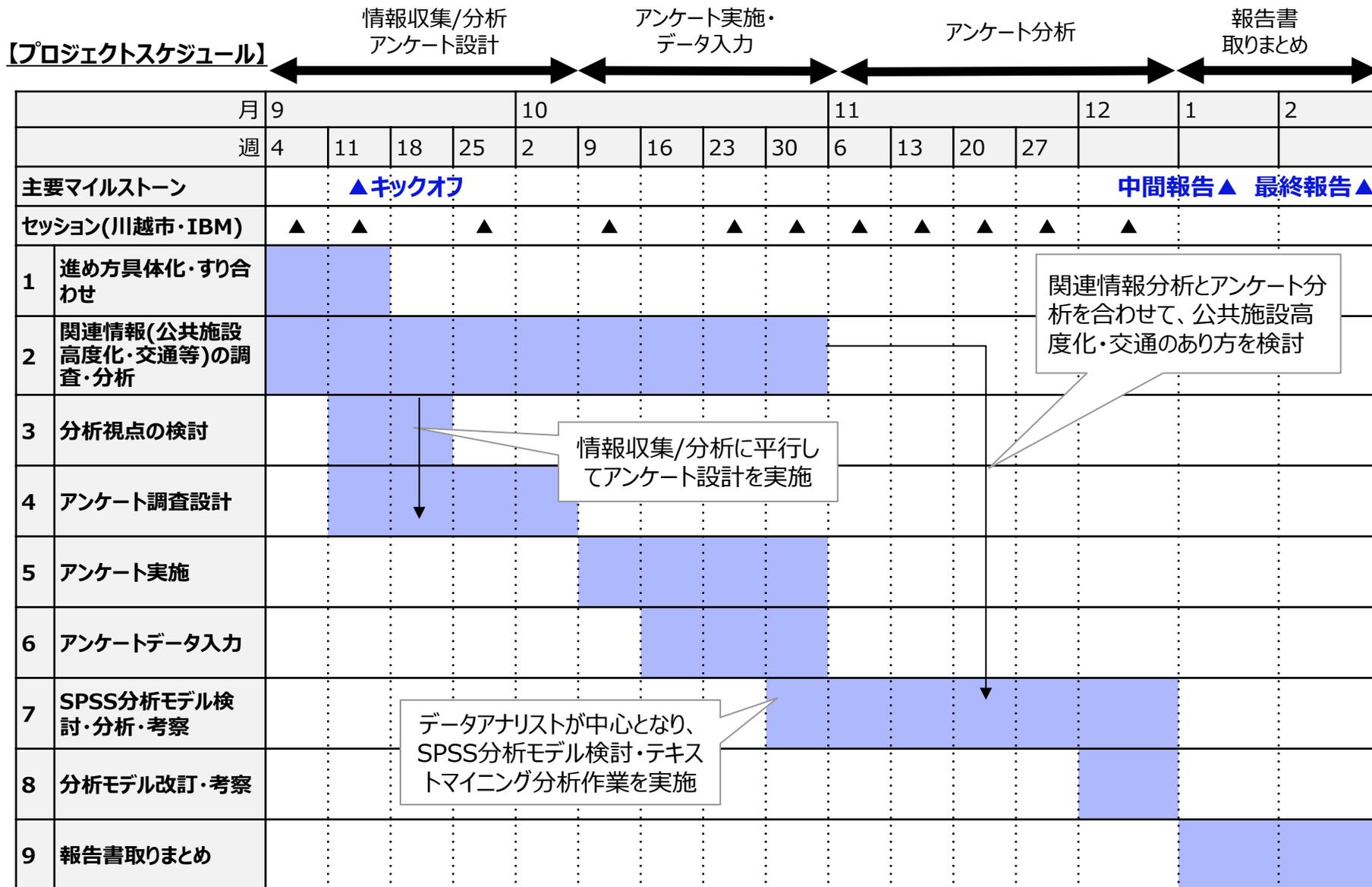
- テキストマイニング機能を活用し、公共施設利用に対する「利用者の生の声」から示唆を導き出し、検討に活かしている。
- データアナリスト(川越市職員)が主体となって分析を実施する事で、今後の政策決定におけるデータ活用に向けた人材の育成および活用事例が得られている。
- 上記を含め、他展開が可能な「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」が抽出できている。

1. 本プロジェクトにおける前提
2. プロジェクトの概要と目的
  1. プロジェクト概要
  2. 目的とゴール
3. プロジェクトのアプローチ
  1. アプローチ
  2. スケジュール
  3. 体制・役割分担
4. 公共施設高度化検討におけるデータ活用検討のポイント
  1. 全体像
  2. データ活用検討のポイント
    1. 分析視点・対象施設の検討
    2. 利用者アンケート調査・設計
    3. 調査実施・データ化作業
    4. SPSSを用いた統計解析・テキストマイニング
    5. データアナリストへのトレーニング
    6. 集約対象候補選定・コストシミュレーション
5. 自治体におけるデータ活用促進に向けた考慮点・気づき
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題
6. 参考資料
  1. 別紙1：公共施設アンケート調査票
  2. 別紙2：川越市公民館利用者アンケート調査結果
  3. 別紙3：SPSSトレーニングセッション資料
  4. 別紙4：コストシミュレーションツールの内容と分析結果

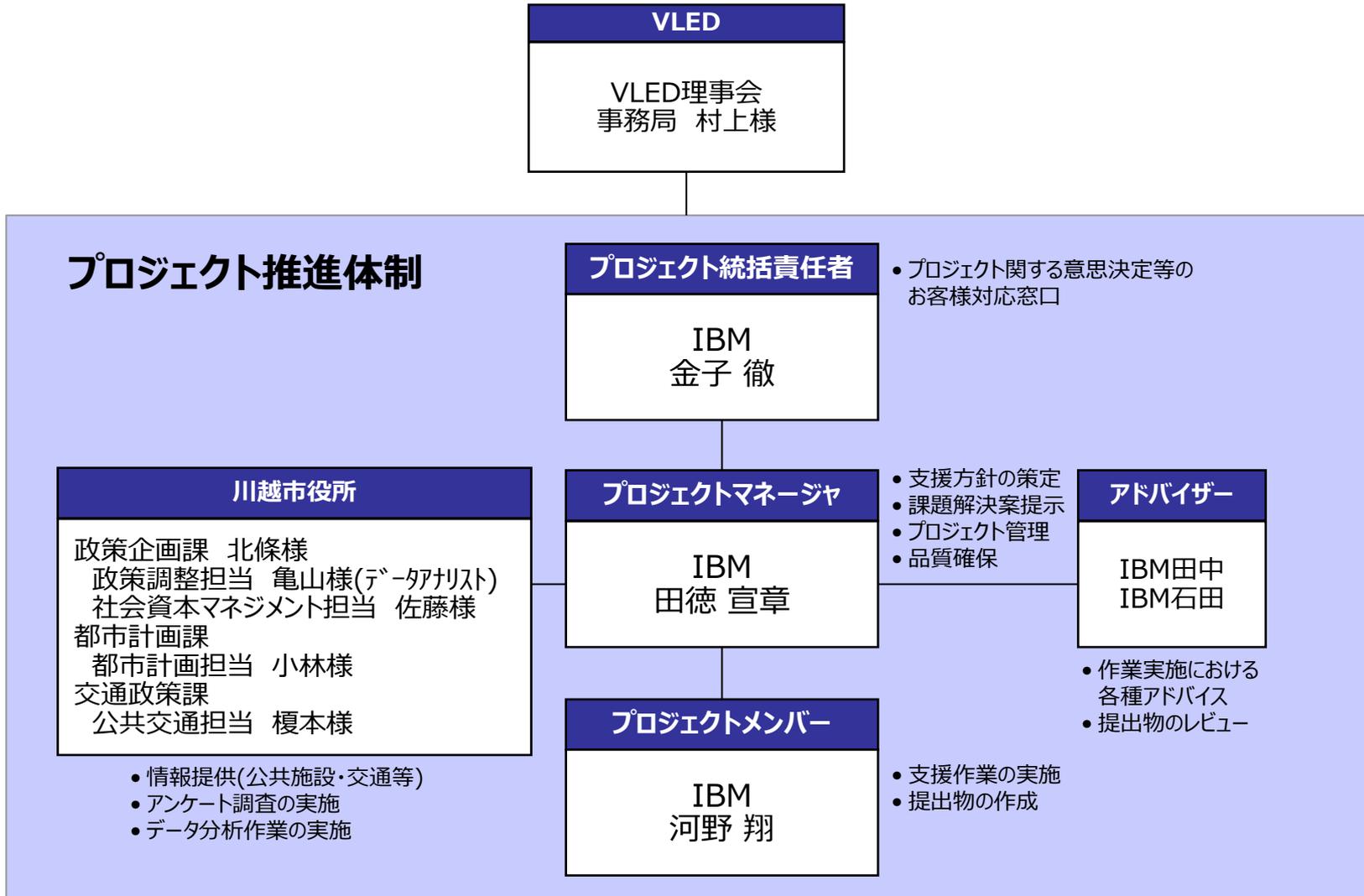
- 本プロジェクトのアプローチは以下のとおり。



- 本プロジェクトは2015年9月～2016年3月の約7ヶ月間で実施しました。



- 本プロジェクトの推進主体は日本アイ・ビー・エムが担い、協力自治体として川越市役所に参加頂き、協働して実施しました。

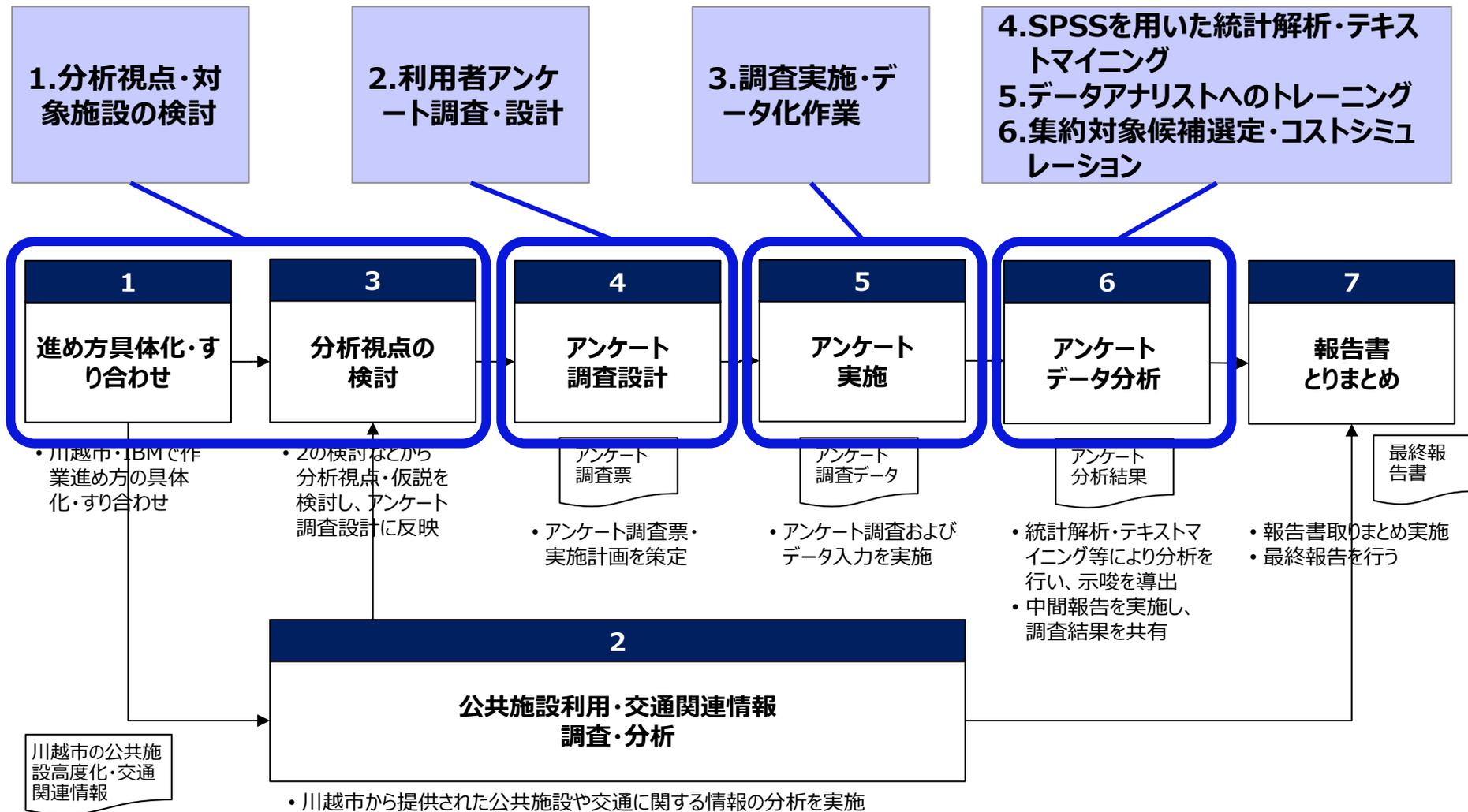


- 本プロジェクトにおける主な役割分担は以下のとおり。

担当	役割分担
	<ul style="list-style-type: none"> <li>中間報告、最終報告会への参加</li> <li>プロジェクトの方向性・分析・報告書に対する評価・助言</li> </ul>
	<ul style="list-style-type: none"> <li>市保有の関連データ提供（公共施設高度化・交通など）</li> <li>セッション準備作業（会場の設定、関係者の日程調整等）</li> <li>アンケート実施協力</li> <li>データ入力作業</li> <li>SPSSを用いたデータ分析作業の実施</li> <li>セッション・討議への参加</li> </ul>
	<ul style="list-style-type: none"> <li>プロジェクト全体推進</li> <li>セッション準備作業（情報収集・整理、討議資料作成）</li> <li>アンケートの調査設計</li> <li>SPSSを用いたデータ分析作業の支援</li> <li>議論のファシリテーション</li> <li>報告書取りまとめ</li> </ul>

1. 本プロジェクトにおける前提
2. プロジェクトの概要と目的
  1. プロジェクト概要
  2. 目的とゴール
3. プロジェクトのアプローチ
  1. アプローチ
  2. スケジュール
  3. 体制・役割分担
4. 公共施設高度化検討におけるデータ活用検討のポイント
  1. 全体像
  2. データ活用検討のポイント
    1. 分析視点・対象施設の検討
    2. 利用者アンケート調査・設計
    3. 調査実施・データ化作業
    4. SPSSを用いた統計解析・テキストマイニング
    5. データアナリストへのトレーニング
    6. 集約対象候補選定・コストシミュレーション
5. 自治体におけるデータ活用促進に向けた考慮点・気づき
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題
6. 参考資料
  1. 別紙1：公共施設アンケート調査票
  2. 別紙2：川越市公民館利用者アンケート調査結果
  3. 別紙3：SPSSトレーニングセッション資料
  4. 別紙4：コストシミュレーションツールの内容と分析結果

- 本プロジェクトを通じて行った工夫や得られた示唆を、以下6つのポイントで示す





- 調査設計では、検討のポイントを網羅した「調査設計フレームワーク」を活用することで、抜け漏れなく進めました。

## <調査設計フレームワーク>

#	フレームワーク	検討内容	検討結果
1	<u>調査の目的</u>	<ul style="list-style-type: none"> <li>調査の目的、および検証仮説を明確化</li> <li>プロジェクト内で共有し、それを踏まえた検討を実施</li> </ul>	<ul style="list-style-type: none"> <li>公共施設利用者アンケートを中心に、公共施設利用状況・属性・満足度の相関等をデータ解析し、討議を通じて、高度化に向けた示唆を得る 【仮説1】利用者の多くが徒歩圏ではなく遠隔地から来ているのではないかな？ 【仮説2】利用者は至近性よりもサービス充実度を求めているのではないかな？</li> </ul>
2	<u>調査対象</u>	<ul style="list-style-type: none"> <li>4つの観点で対象を検討</li> <li>1.将来的に高度化検討対象となりうる施設</li> <li>2.アンケート調査への協力が得られる施設</li> <li>3.当該期間に必要なデータが得られる施設</li> <li>4.交通面の課題が内在しそうな施設</li> </ul>	<ul style="list-style-type: none"> <li>1の観点で確認を行った所、川越市では、高度化・集約化施設対象として具体的に検討が進められている施設は特定されていなかった</li> <li>そこで、本プロジェクトの検討・分析プロセスを実証するため、2～4の観点から、「公民館(17館)」を仮の調査対象として設定した</li> </ul>
3	<u>調査項目</u>	<ul style="list-style-type: none"> <li>「1.利用者属性」・「2.施設の利用状況」・「3.施設に対する評価」・「4.その他自由記述」の観点で調査項目案を導出</li> <li>その後、「目的に照らした優先度」・「回答/分析の容易性」の観点で、選定・絞込みを実施</li> </ul>	<ul style="list-style-type: none"> <li>(参照：別紙1「公民館に関するアンケート調査票」)</li> </ul>
4	<u>調査規模</u>	<ul style="list-style-type: none"> <li>「標本調査法における有限母集団での標本数算出式」を用い、理想サンプル数を算出</li> <li>予算や期間などリソース制約および求める信頼性に鑑み、必要に応じて調整</li> </ul>	<ul style="list-style-type: none"> <li>アンケート配布枚数：1000枚</li> <li>目標有効サンプル数：400件 (目安として1施設20-25件程度、バリエーションのため4グループ/個人以上の対象者から収集)</li> </ul>
5	<u>調査期間</u>	<ul style="list-style-type: none"> <li>調査・分析に必要な期間、対象施設の都合等も考慮し、調査期間を設定</li> </ul>	<ul style="list-style-type: none"> <li>10/10(土)配布開始～2週間(目標有効サンプル数収集に依存)</li> </ul>
6	<u>調査方法</u>	<ul style="list-style-type: none"> <li>様々な調査方法の長所・短所を比較し決定</li> </ul>	<ul style="list-style-type: none"> <li>公民館利用申請時に、申請者に複数枚配布し記入依頼</li> </ul>
7	<u>解析方法</u>	<ul style="list-style-type: none"> <li>SPSSを用いた統計解析・テキストマイニングにより分析を実施</li> <li>作業プロセスごとに分け、役割分担・実施方法を検討</li> </ul>	<ul style="list-style-type: none"> <li>【調査設計】優先度・回答し易さなどを考慮し、1枚裏表程度に設問絞込み</li> <li>【印刷】IBMにて印刷し、川越市役所に郵送</li> <li>【配布・回収】各公民館にて配布・収集</li> <li>【データ化作業】市にてデータ化実施</li> <li>【分析作業】市・IBMにて、SPSSを用いた統計解析・テキストマイニング</li> </ul>
8	<u>日程</u>	<ul style="list-style-type: none"> <li>調査スケジュールおよびプロジェクト全体スケジュールに鑑み、必要に応じて調整</li> </ul>	<ul style="list-style-type: none"> <li>(参照：本資料P8「スケジュール」)</li> </ul>

### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。



- 検討に先立ち、本調査の目的および仮説をメンバー間での共有を図りました。**序盤で意識を一つにする**事で、以降の対象施設・調査項目検討や分析作業の効果的な推進に繋がりました。

## 調査の目的

- 川越市の公共施設利用者アンケートを中心としたデータ解析を通じて、公共施設利用状況・利用者属性・施設満足度の相関等を分析し、公共施設高度化に向けた示唆を得る

### (仮説 1)

分散設置施設の利用者の相当数が徒歩圏ではなく遠隔地から来ているのではないか？

- ✓ 利用者の相当数が自家用車または公共交通機関(コミュニティバスを含む)で来訪しているのではないか？
- ✓ 利用者で最寄の施設ではなく遠方の施設を利用している人が一定数いるのではないか？

### (仮説 2)

分散設置施設の利用者は施設の至近性よりもサービス充実度を求めているのではないか？

- ✓ 提供サービスに対する利用者の満足度が高い施設においては、市外からの利用者が相当数いるのではないか？
- ✓ 提供サービスに対する利用者の満足度が低い施設(の地域)においては、他施設(域外や市外)への流出が相当数いるのではないか？

(仮説3) 公共施設利用者のアクセス手段としてのコミュニティバス等公共機関を最適化することで、公共施設の集約化とサービス高度化が可能となるのではないか？



- 調査対象の検討では、まず対象施設を一覧化したうえで、プロジェクト目的に照らし**4つの観点**で整理を行いました。
- 現時点、具体的に高度化・集約化対象となっている施設があるかを確認した所、**市からは「具体的検討が進められている施設はない」と**の回答がありました。
- 本プロジェクトでの調査対象として仮の対象施設を設定する必要があったため、市担当者・施設を所管する原課等との協議を経て、**本プロジェクトに限ってのあくまで仮の調査対象として「公民館(17館)」**と設定しました。

## ①市が保有する公共施設の 情報収集を実施

- ・ 市民センター・連絡所
  - ・ 公民館
  - ・ 図書館
  - ・ 美術館
  - ・ 市立博物館
  - ・ 生活・文化施設
  - ・ 教育機関
  - ・ 保育施設
  - ・ 福祉施設
  - ・ 医療機関
  - ・ 運動・公園施設
  - ・ 消防署 など
- ※川越市公共施設マネジメント白書(平成25年3月)等より

## ②プロジェクト目的に照らし、 4つの観点で整理

- 1.市として将来的に高度化検討対象となる施設
- 2.アンケート調査への協力が得られそうな施設
- 3.調査期間中に必要なデータ量が得られそうな施設
- 4.交通面の課題が内在しそうな施設

## ③施設を所管する 原課と協議

- ・川越市として、高度化の**具体的検討が進められている施設は、ない**

## ④総合的に判断し、 調査対象を設定

- (本プロジェクトに限っての、あくまで仮の調査対象として)
- ・ **公民館(17館)**

### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。



- また、川越市から本プロジェクトに関して提供可能な情報の内容確認を行い、公民館利用の分析に有用と考えられるデータの絞り込みを行いました。

原課	データ名称	公民館との関連	媒体
産業振興課	中心市街地における歩行者・自転車の交通量	×	Excel
観光課	観光アンケートデータ	×	Excel
	観光案内所利用者数（月別、目的別:観光・宿泊・スポーツ・生活）※川越・本川越・仲町	×	Excel
	市役所駐車場データ（土日、台数）	×	Excel
	まつり会館入館者数（月毎、国内外の2分類）	×	Excel
	入館者数（月毎、国内外の2分類） ※博物館、美術館、本丸御殿、蔵造り資料館	×	Excel
交通政策課	コミュニティサイクル	×	Excel
	一番街交通量調査（一方通行化社会実験時）	×	Excel
	川越市公共施設マネジメント白書（平成25年版）	○	PDF
	川越シャトルOD調査	○	PDF
	路線バスダイヤ最適化システムの活用による川越シャトル検証調査報告書（イーグルバス社提供）	○	Excel,PPT (CDR)
	市内循環バス「川越シャトル」利用者アンケート集計結果（2013年3月）	○	Word
高齢者いきがい課	後楽会館ののべ利用人数	×	Excel
農業ふれあいセンタ	センター利用者データ	×	Excel
情報統計課	統計かわごえ（※施設利用者数・コスト情報等）	○	Excel

## 【本プロジェクトにおける前提】

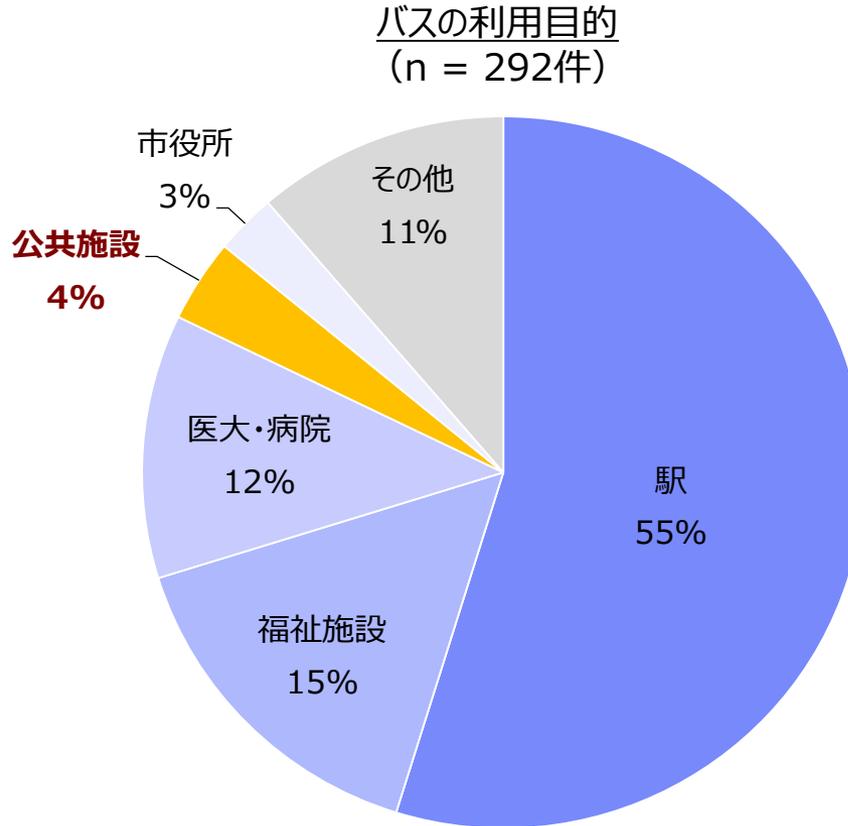
- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。

- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

- 公共交通機関に係る関連既存調査は、「①川越シャトルOD調査」・「②路線バスダイヤ最適化システムの活用による川越シャトル検証調査報告書」・「③市内循環バス「川越シャトル」利用者アンケート集計結果」の3件。
- そのうち、①②は乗降地を把握できますが利用目的が分からないため、公民館利用との関連付けが難しい状態。他方③は項目に「**利用目的**」があったため、それを用いて、公共交通機関と公民館利用との関連を分析しました。

調査名	①川越シャトルOD調査	②路線バスダイヤ最適化システムの活用による川越シャトル検証調査報告書	③市内循環バス「川越シャトル」利用者アンケート集計結果(2013年3月)
データ概要	<ul style="list-style-type: none"> <li>乗降者人数(系統別・バス停別)</li> <li>乗車中人数(系統別)</li> <li>区間乗車人数(系統別・便別)</li> <li>系統別OD情報(系統別)</li> <li>利用者属性(券種から推定)</li> <li>支払い方法</li> </ul>	<ul style="list-style-type: none"> <li>乗降者人数(バス停別)</li> <li>乗車中人数(ダイヤ別・バス停別)</li> <li>OD情報(ダイヤ別)</li> <li>遅延箇所(ダイヤ別・バス停別)</li> <li>支払い方法(ICカードのみ集計)</li> </ul>	<ul style="list-style-type: none"> <li>利用者アンケート                             <ul style="list-style-type: none"> <li>利用者属性</li> <li>利用頻度</li> <li><b>利用目的</b></li> <li>前後の鉄道利用</li> <li>自動車の保有</li> <li>自由記述等)</li> </ul> </li> </ul>
提供者	<ul style="list-style-type: none"> <li>川越市</li> </ul>	<ul style="list-style-type: none"> <li>イーグルバス株式会社</li> </ul>	<ul style="list-style-type: none"> <li>川越市(公共交通利用促進協議会)</li> </ul>
調査期間	<ul style="list-style-type: none"> <li>2015/6/24(水)～6/30(火)</li> </ul>	<ul style="list-style-type: none"> <li>2012/11/1(木)～12/31(月)</li> </ul>	<ul style="list-style-type: none"> <li>2013/1/21(月)～1/28(月)</li> </ul>
調査対象	<ul style="list-style-type: none"> <li>川越シャトルの全運行ダイヤ(平日140便/日、休日110便/日)</li> </ul>	<ul style="list-style-type: none"> <li>イーグルバス対象ダイヤ</li> </ul>	<ul style="list-style-type: none"> <li>全系統のバス利用者</li> <li>車内に配布ボックスを設置して利用者に持ち帰ってもらい、郵送で川越市役所へ回収(料金受取人払い)</li> </ul>
データ数	<ul style="list-style-type: none"> <li>取得乗車人数：7,885人分</li> </ul>	<ul style="list-style-type: none"> <li>所得乗車人数：70,014人分(日平均値から推計)</li> </ul>	<ul style="list-style-type: none"> <li>回収部数：174部(車内準備部数：980部=70部×14台)</li> </ul>

- 「③市内循環バス「川越シャトル」利用者アンケート集計結果」を基に、バスの利用目的の分析を行った所、**公共施設(公民館、図書館など)を選択したのは4%程度**に留まっており、川越市では**公民館を目的とした公共交通機関利用は活発ではない**、と考えられます。



※市内循環バス「川越シャトル」利用者アンケート集計結果（2013年3月）より作成

【本プロジェクトにおける前提】

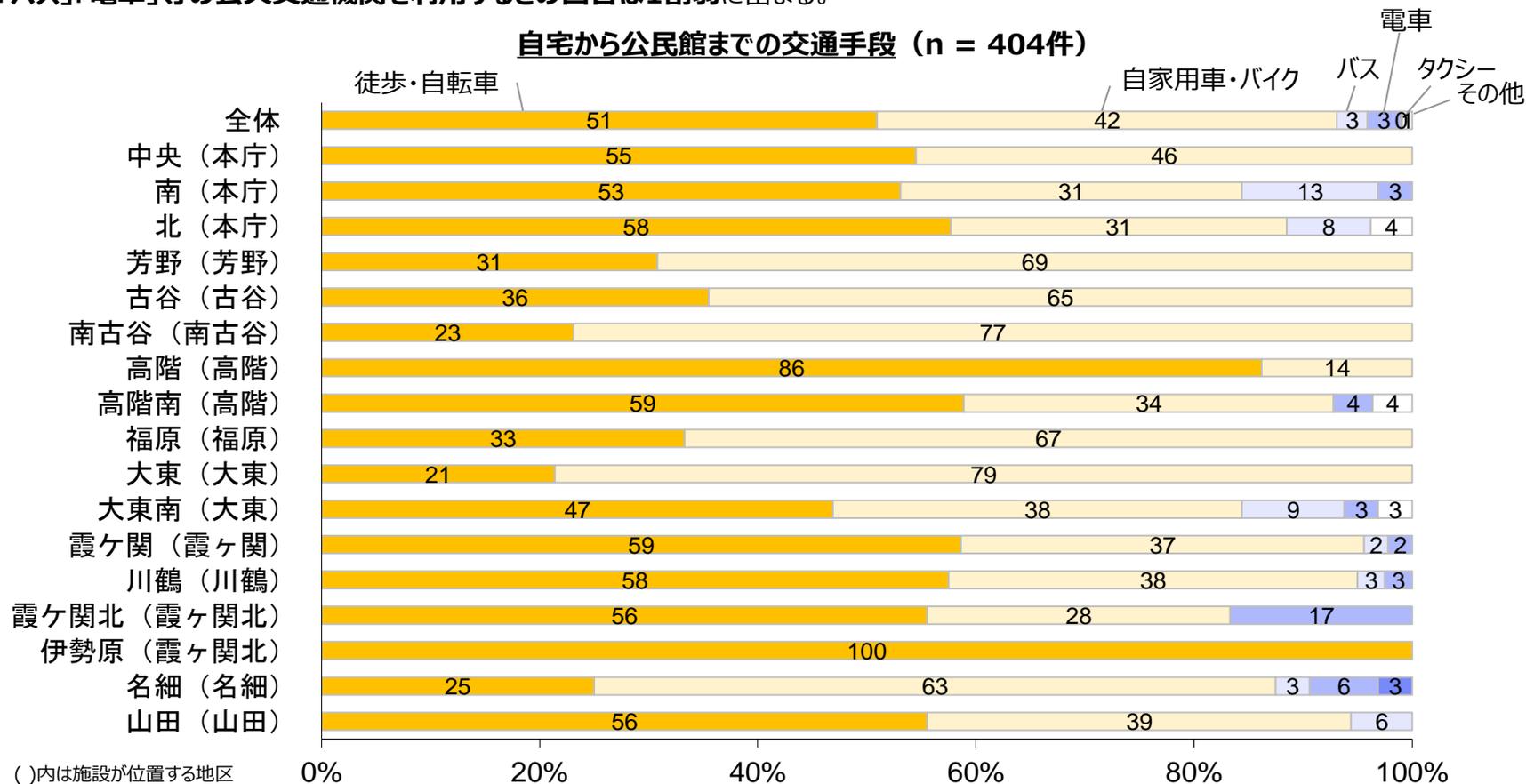
- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

- 今回行った公民館利用者アンケート結果でもバス・電車等の公共交通機関利用者は1割弱に留まっていたことから、市と協議の上、高度化における影響は限定的と判断し、既存交通データの検討優先度を落としました。

**【アクセス | 自宅から公民館までの交通手段】 公共交通機関を利用している人は少ない。徒歩に加え、自家用車・バイクが多い**

- 5割が「徒歩・自転車」、4割が「自家用車・バイク」を利用していると回答。
- 「バス」「電車」等の公共交通機関を利用するとの回答は1割弱に留まる。

自宅から公民館までの交通手段 (n = 404件)



\*1 : ( )内は施設が位置する地区  
\*2 : 回答無し76件は母数から除く

**【本プロジェクトにおける前提】**

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。  
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

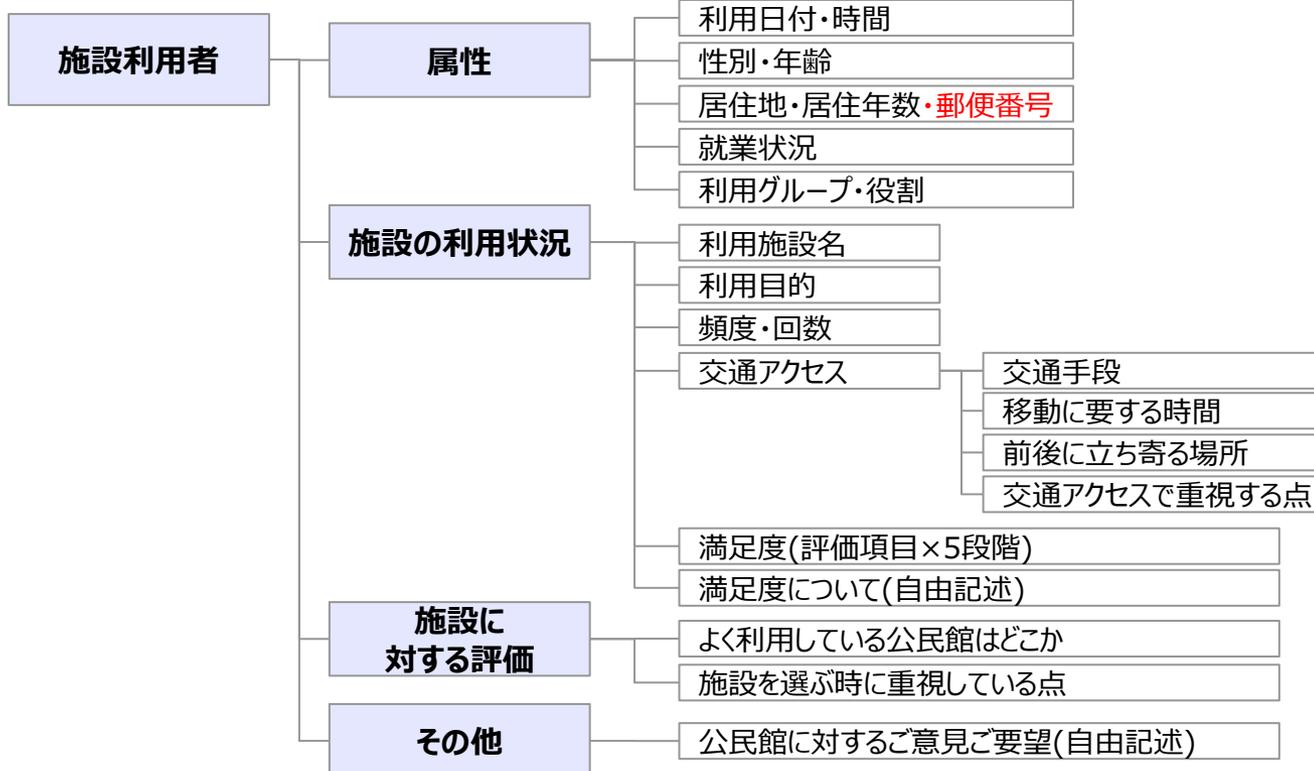


- アンケート調査票の項目検討は、以下3ステップで実施しました
- 調査項目の骨子として4つの分類を設定し、**参考事例・ブレインストーミングにより調査項目の洗出しを実施。目的に照らした優先度・回答者の負担・回答率・データ化作業量等を考慮し、A4裏表1枚(13問)への絞込みを行いました。**
- 利用者と公共施設の位置関係を把握するために、利用者属性調査項目として「郵便番号」を設定しました。

### ① 調査項目の骨子を定義

### ② 必要・有用と考えられる調査項目を列挙

### ③ 目的に照らした優先度・回答のし易さを考慮し、設問の絞込みを実施



利用者アンケート調査票 (A4裏表1枚)



- また、実際の調査票を作成するにあたっては、回答者の回答しやすさ、集計・分析のしやすさなどの観点を意識し、帳票の構成や設問を工夫・配慮しました。

### 【帳票番号】

- 配布地域ごとに番号帯を設定
- データ化時に視認しやすい位置に配置
- 差込み印刷で採番

### 【郵便番号】

- より正確な住所情報を取得する目的
- 公共施設までの距離算出に活用

### 【A4裏表1枚に集約】

- 回答者の負担・回答率を考慮し、A4裏表1枚に収めた

No. XXXX

### 公民館に関するアンケート

市では、公民館の利用状況について調査しています。今後の運営の参考にさせていただきたいので、アンケートにご協力をお願いします。

①利用日 平成27年 月 日 (午前・午後・夕方)  
 ②性別 ア：男性 イ：女性  
 ③年齢 ア：20歳未満 イ：20代 ウ：30代 エ：40代 オ：50代 カ：60代 キ：70歳以上  
 ④住まい ア：本庁 イ：芳野 ウ：古谷 エ：南古谷 オ：高階 カ：福原 キ：大東 ク：霧ヶ関  
 ケ：川鶴 コ：霧ヶ関北 サ：名細 シ：山田 ス：市外(自治体名： )  
 ⑤郵便番号 差支えなければ郵便番号をご記入ください。(〒 - )  
 ※住所を記入する必要はありません。  
 ⑥居住年数 ア：1年未満 イ：1~3年未満 ウ：3~5年未満 エ：5~10年未満  
 オ：10年以上 カ：分からない

問1. 本日はこの公民館をどのような目的で利用していますか。当てはまるものをすべて選んで○を付けてください。(複数回答)

1 クラブ・サークル活動	2 講座等の受講	3 自治会等の地域活動
4 勉強・学習のため	5 その他( )	

問2. 直近1年間のこの施設の利用頻度について当てはまるもの1つに○を付けてください。

1 週に2~3回以上	2 週1回	3 月に2~3回
4 月に1回以下	5 その他( )	

問3. あなたは次のどれに当たりますか。当てはまるもの1つに○を付けてください。

1 一般団体利用の予約者	2 一般団体利用のメンバー(予約者以外)	3 公民館登録団体利用の予約者
4 公民館登録団体利用のメンバー(予約者以外)	5 個人の利用者	6 その他( )

問4. (問3で、1~4とお答えの方) 本日は何人ぐらいのグループで利用していますか。当てはまるもの1つに○を付けてください。

1 2~5人	2 6~10人	3 11~20人
4 20人以上	5 分からない	

問5. よく利用する公民館はどこですか。次の中から多し順に番号(1~3)を記載してください。(3つまで)

中央公民館	南公民館 (ウェス川域内)	北公民館
芳野公民館	古谷公民館	南古谷公民館
高階公民館	高階南公民館	福原公民館
大東公民館	大東南公民館	霧ヶ関公民館
川鶴公民館	霧ヶ関北公民館	伊勢原公民館
名細公民館	山田公民館	市外の公民館 (名称： )
利用していない		

表

### 【立ち寄り施設】

- 利用者動線・他施設との関連を調査することで、交通施策や複合施設化への示唆を狙った

裏

問6. 公民館を利用するときに重視することは何ですか。当てはまるものを3つ選んで○を付けてください。(複数回答)

1 予約が取りやすい	2 利用料金が安い	3 立地・交通アクセスが良い
4 使える曜日や時間が活動と合う	5 設備が充実している(会議室の数・広さ等)	6 十分な駐車スペースがある
7 職員の対応が良い	8 その他( )	

問7. この公民館の満足度を教えてください。①~⑦について、5段階評価で当てはまるものを選んで○を付けてください。

	1. とても良い	2. 良い	3. 普通	4. 悪い	5. とても悪い
--	----------	-------	-------	-------	----------

(右記のように○を付けてください) →

①予約が取りやすい	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
②利用料金が安い	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
③立地・交通アクセスが良い	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
④使える曜日や時間が活動と合う	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
⑤設備が充実している	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
⑥十分な駐車スペースがある	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
⑦職員の対応が良い	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

問8. 問7で評価した理由を教えてください。(すべての項目についてお答えいただく必要はありません)(自由記述)

問9. ご自宅からこの公民館までの交通手段は何ですか。当てはまるものをすべて選んで○を付けてください。(複数回答)

1 徒歩・自転車	2 自家用車・バイク	3 バス
4 電車	5 タクシー	6 その他( )

問10. ご自宅からこの公民館までの移動にかかる時間はどれくらいですか。当てはまるもの1つに○を付けてください。

1 15分未満	2 15~30分未満	3 30分~1時間未満
4 1時間以上		

問11. 公民館までのアクセスについて重視する点で当てはまるものをすべて選んで○を付けてください。(複数回答)

1 自宅から近いこと	2 駅から近いこと	3 バス停から近いこと
4 商業施設・商店街から近いこと	5 他の公共施設から近いこと	6 その他( )

問12. 公民館を利用する前後に立ち寄る場所はどこですか。当てはまる場所をすべて選んで○を付けてください。

1 スーパー・コンビニ	2 飲食・喫茶	3 病院
4 郵便・金融機関	5 公共施設	6 学校・仕事
7 保育園・幼稚園	8 自宅または友人・知人宅	9 その他( )

問13. その他、公民館に対するご意見・ご要望がありましたら、ご自由にご記入ください。

本アンケートは「川越市 政策企画課 (TEL : 049-224-5503)」までお問い合わせください。

### 【自由記述欄の枠】

- 自由記述で得たい情報量と、データ化作業負担を考慮し、経験も踏まえてスペースを設定

参考：別紙1「公民館に関するアンケート調査票」

### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。



- 調査規模の検討では、「**標本調査法における有限母集団での標本数算出式**」を用いて、必要な標本数を**383.9人**と算出。回収率を40%程度と想定し、調査票の**配布数を1,000枚に決定**しました。

### 【標本調査法における有限母集団での標本数算出式】

$$\text{標本数 (人)} = \frac{\text{(標準誤差)}}{\text{(信頼水準)}} \times \frac{\text{(標準誤差)}}{\text{(信頼水準)}} \times \frac{\text{(母数)} - 1}{\text{(母比率)} * (1 - \text{母比率})} + 1$$

母集団

「回答比率」とは、調査対象者の回答比率のこと。参考情報がない場合は0.5を設定  
 「標準誤差」とは、容認できる誤差の範囲のこと。例えば5%に抑えたい場合は0.05と設定  
 「信頼水準」とは、正しく判断できる確率のこと。国が行う標本調査では95%(1.96)と設定

- 公民館(17館)の年間利用者数を分母として計算を行うと、必要な標本数は**383.9人**  
 - 母集団：公民館(17館)の年間利用者数907,465人(2011年度)、回答比率：0.5、標準誤差0.05、信頼水準95%(1.96)

$$\text{標本数 (人)} = \frac{907,465}{\frac{0.05}{1.96} \times \frac{0.05}{1.96} \times \frac{(907,465) - 1}{(0.5) * (1 - 0.5)}} + 1$$

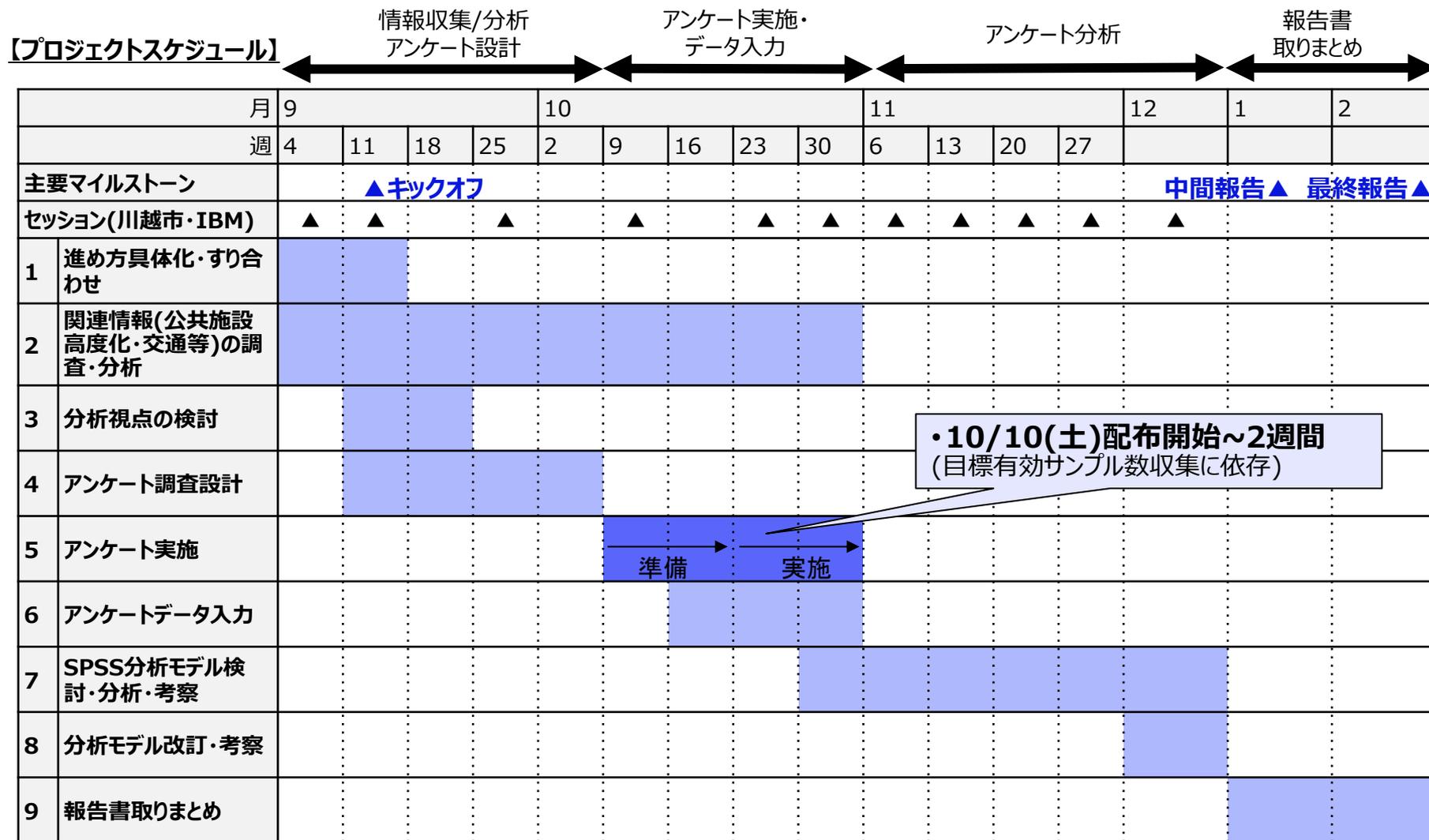
• **本調査では、「アンケート配布枚数：1000枚」、「目標有効サンプル数：400件」を目標設定**

**【本プロジェクトにおける前提】**  
 - 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。  
 - 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

## 2.利用者アンケート調査・設計 調査期間



- 調査期間の検討では、調査・分析に必要な期間、対象施設の都合、市アンケート調査の経験値等を考慮し、400件の有効回答回収に掛かる期間を「**10/10(土)配布開始～2週間**」と設定。それに先立って帳票印刷・配布、施設担当者への説明等の準備を実施しました



- 調査方法では、「回収率の向上」・「データバリエーションの確保」の2つを重視し、本プロジェクトの特徴・条件と、各調査手段を比較し、アンケート(会場調査)を選択しました。

調査手段		長所	短所
<b>面接調査</b>	対面で聞き取りを実施	<ul style="list-style-type: none"> <li>質問の内容・量といった制限が少ない</li> <li>対面で依頼・質問することで、回収率改善や質の高い回答が得やすい</li> <li>確実に対象者に調査することができる</li> </ul>	<ul style="list-style-type: none"> <li>人件費・会場費・交通費等がかかる</li> <li>調査者の質により、回答の誘導が起きるリスク</li> <li>調査者の教育・トレーニング工数が必要</li> <li>在宅率低下・オートロック等による障害</li> </ul>
<b>電話調査</b>	電話で聞き取りを実施	<ul style="list-style-type: none"> <li>対象者が地理的に分散していても、効率よく調査でき、回答も即時得られる</li> <li>面接調査と比較し、調査期間・費用を抑制できる</li> <li>確実に対象者に調査することができる</li> </ul>	<ul style="list-style-type: none"> <li>相手の時間を直接的に拘束するため、質問量はそれに見合う量に制限される</li> <li>内容も、音声だけで理解できるものに制限される</li> <li>電話帳不掲載を希望する世帯も多く、回答の代表性にも留意が必要</li> </ul>
<b>アンケート(郵送)</b>	調査票を対象者に郵送	<ul style="list-style-type: none"> <li>調査員を使わないため、比較的が安価</li> <li>対象者が地理的に分散していても、効率よく調査できる</li> </ul>	<ul style="list-style-type: none"> <li>対面調査に比べ、回収率が低く、予想もしづらい</li> <li>調査票の印刷・配布・回収などの手間・コストや、データ入力作業に工数がかかる</li> <li>誤入力・回答漏れのリスク</li> <li>本人確認が難しい・周囲の影響を受ける可能性</li> </ul>
<b>アンケート(会場調査)</b>	調査票を会場を対象者に渡し、回答	<ul style="list-style-type: none"> <li>アンケート(Web)と同様の長所</li> <li>対面で依頼することで、回収率改善や質の高い回答が得やすい</li> </ul>	<ul style="list-style-type: none"> <li>アンケート(Web)と同様の短所</li> <li>調査期間中に会場に訪れた人が対象となるため、回答の代表性に留意が必要</li> </ul>
<b>アンケート(Web)</b>	インターネットを用いてアンケートを実施	<ul style="list-style-type: none"> <li>調査票の配布・回収や、データ入力といった手間が省けるため、調査期間を抑えることができる</li> <li>入力制御により、誤入力・回答漏れを防止できる</li> </ul>	<ul style="list-style-type: none"> <li>インターネット環境が整っていない人や場所では答えることができないため、回答者が「パソコン利用者」「携帯利用者」「若年層」に偏りがちななど、回答の代表性に留意が必要</li> </ul>
<b>ソーシャルメディア分析</b>	SNS等に蓄積されたデータを分析	<ul style="list-style-type: none"> <li>注目度・認知度が高いテーマの場合、大量のデータを得やすい</li> </ul>	<ul style="list-style-type: none"> <li>アンケート(Web)と同様の短所</li> <li>雑多かつ膨大なデータの分析には、分析ツールなどの整備・利用コストが必要</li> <li>注目度・認知度の低いテーマの場合、十分なデータ量が集められない</li> </ul>



- アンケート(会場調査)は、**調査期間中に施設を利用した利用者・グループを対象に手渡し・同日回収**で実施。
- また、バリエーション確保のため、**施設ごとのデータ数設定、1グループあたりの配布制限、配布タイミングをなるべく広範に取る**等の工夫を行いました。
- また、必要標本数確保できるよう、実施現場と回収状況をこまめに確認し、必要に応じ調査期間延長も視野に準備を行いました。(実際は当初予定期間で十分な標本数を確保)

<b>対象者</b>	<ul style="list-style-type: none"> <li>調査期間中に公民館施設を利用した利用者・利用者グループ</li> </ul>
<b>配布方法</b>	<ul style="list-style-type: none"> <li>利用申請時に、職員から申請者に調査票を複数枚(4枚/グループ)手渡して記入依頼。利用終了の際に同日回収を実施</li> <li>「午後2」枠の終了時には職員が退勤しているため、アンケート投入ボックスを設置</li> <li>バリエーション確保のため、目安として「1施設あたり20-25件」、「4グループ以上の対象者から収集」と目標設定</li> </ul>
<b>配布タイミング</b>	<ul style="list-style-type: none"> <li>施設予約枠の4枠のうち、<b>職員が配布可能な3枠を対象に、全ての曜日で配布実施</b>することで、データバリエーションを確保</li> </ul>

公民館予約枠	月	火	水	木	金	土	日
【午前】午前9時から正午	○	○	○	○	○	○	○
【午後1】午後1時から3時30分	○	○	○	○	○	○	○
【午後2】午後4時から6時30分	○	○	○	○	○	○	○
【夜間】午後7時から9時30分	/	/	/	/	/	/	/

**【本プロジェクトにおける前提】**

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

### 3.調査実施・データ化作業 調査における役割分担

- 作業プロセスごとに分けて、川越市・IBMの役割分担を検討。市はアンケート調査・データ入力・分析作業など、調査分析の主体を担当し、IBMは調査設計支援・分析作業支援などを担当しました。
- 「データ化作業」は、個人情報保護の観点から、特に自治体にて行うべき業務として位置付け、実施しました。

#### 【調査における役割分担】

個人情報保護の観点で、特に自治体にて行うべき業務

凡例

主担当

	調査設計	アンケート調査票印刷	アンケート調査実施	データ化作業	分析作業	報告書作成
 <b>川越市</b>	調査設計	—	公共施設にて実施・回収	市にて実施	分析作業	報告書作成 確認・助言
	調査設計支援	印刷	—	データ入力用フォーマット検討	分析作業支援	報告書作成支援

各プロセスにおける業務支援を担当

### 3.調査実施・データ化作業 データ入力用フォーマットによる作業効率化



- データ入力作業を効率的に進めるため、分析時に利用する解析ツール（SPSS）での取り込みを想定したデータ入力用フォーマットを作成しました。
- 作業者の意見を貰いつつ、項目ごとに選択肢の一覧や記入例を追加等、作業を行いやすいよう工夫を実施したことで、自由記述欄も含めた全データ入力作業を効率的に進める事ができ、想定期間内に完了させることができました。

#### 【データ入力用フォーマット】

No. 0 (属性情報)							
設問No.	①-1	①-2	②	③	④	④-FA	⑤
設問内容	利用日	時間帯	性別	年齢	住まい	市外(自治体名)に記述がある場合	郵便番号
(複数選択の場合) 選択番号							
(複数選択の場合) 選択内容							
入力時の選択肢	何月何日	0:未記入 1:午前 2:午後 3:夕方	ア:男性 イ:女性 ウ:未記入	ア:20歳未満 イ:20代 ウ:30代 エ:40代 オ:50代 カ:60代 キ:70代 ク:未記入	ア:本庁 イ:芳野 ウ:古谷 エ:南古谷 オ:高階 カ:福原 キ:大東 ク:霧ヶ関 ケ:川鶴 コ:霧ヶ関北 サ:名細 シ:山田 ス:市外 セ:未記入	自由記述	自由記述
(記入例)	10月8日	1	ア	カ	ス	さいたま市	330-0061
0000							
0001							
0002							
0003							
0004							
0005							

①入力用フォーマット準備  
・あらかじめツール(SPSS)での取り込みを前提とした構成でフォーマットを準備し、入力精度を向上

②凡例の追加  
・作業者が入力作業を行いやすいよう、項目ごとに選択肢の一覧や記入例を追加

③凡例の追加  
・選択項目の場合は、リスト昨日を活用し、誤入力を抑止・作業効率向上

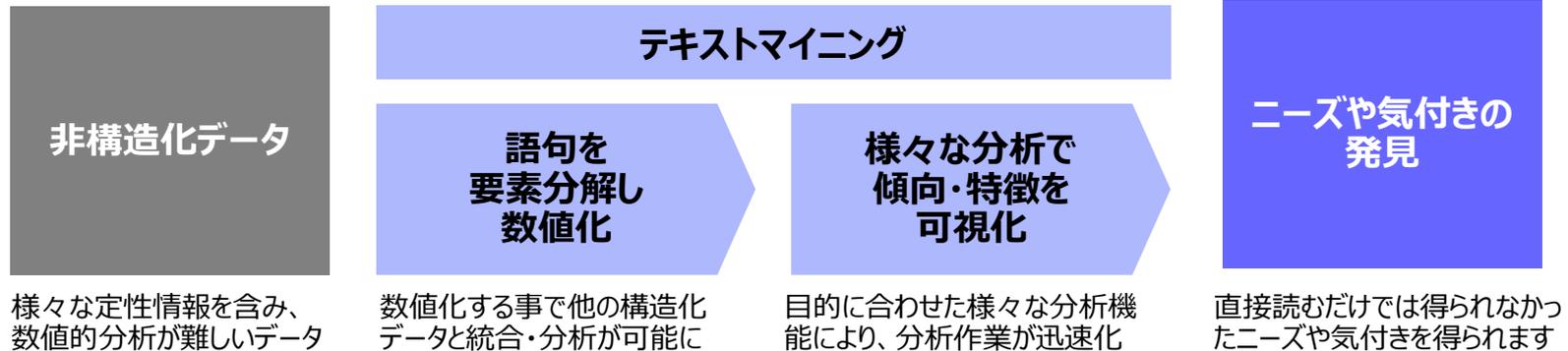
参考：別紙 VLED\_アンケート入力フォーマット\_151005\_v1.1.xlsx

#### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

1. 本プロジェクトにおける前提
2. プロジェクトの概要と目的
  1. プロジェクト概要
  2. 目的とゴール
3. プロジェクトのアプローチ
  1. アプローチ
  2. スケジュール
  3. 体制・役割分担
4. 公共施設高度化検討におけるデータ活用検討のポイント
  1. 全体像
  2. データ活用検討のポイント
    1. 分析視点・対象施設の検討
    2. 利用者アンケート調査・設計
    3. 調査実施・データ化作業
    4. SPSSを用いた統計解析・テキストマイニング
    5. データアナリストへのトレーニング
    6. 集約対象候補選定・コストシミュレーション
5. 自治体におけるデータ活用促進に向けた考慮点・気づき
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題
6. 参考資料
  1. 別紙1：公共施設アンケート調査票
  2. 別紙2：川越市公民館利用者アンケート調査結果
  3. 別紙3：SPSSトレーニングセッション資料
  4. 別紙4：コストシミュレーションツールの内容と分析結果

- テキストマイニングとはアンケートの自由記述欄等で寄せられる「非構造化データ」を、客観的・効率的に分析する手法。大量のテキスト情報から「ニーズ」や「気付き」を発見することができ、企業のマーケティング等で実績を上げています。



アンケートの自由記述  
テキストやインタビュー

区分	街のイメージ
#	628
1	街 (89)
2	若者 (39)
3	ある (35)
4	雑多 (34)
5	活気 (28)
6	最先端 (26)

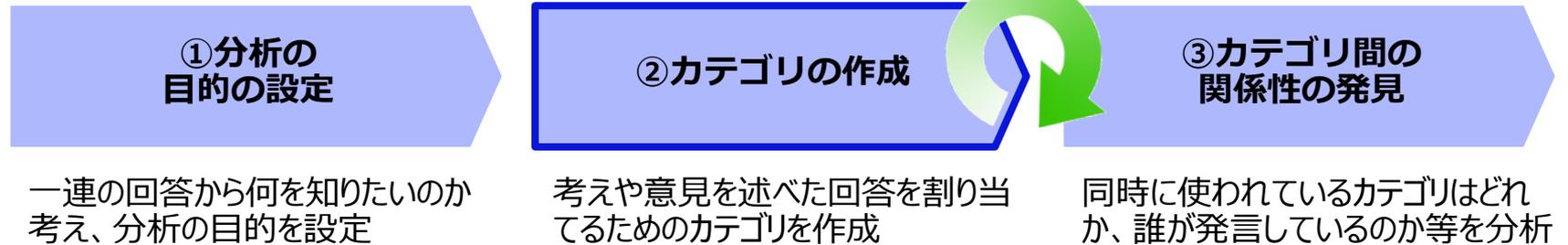


### 〈テキストマイニングの有効性〉

- ・大量のテキストデータの集計・分析の効率化
- ・手作業では抽出できない相関・傾向・特徴の発見
- ・数値化によって他の構造化データと統合分析可能
- ・裏付けのある評価、客観性の確保が可能

- テキストマイニングは「①分析の目的の設定」「②カテゴリの作成」「③カテゴリ間の関係性の発見」の順番で行います。
- このうち、分析に適した“カテゴリの作成”が、最も重要な作業となります。カテゴリとは「考えや意見を述べた回答を割り当てるための“箱”(≒似た意味を持つ語句のかたまり)」のことです。
- **カテゴリ導出にあたって定型的・汎用的な検討方法はなく、ケースに合わせた試行錯誤・議論を重ね、精度を上げる必要があります。**

試行錯誤・議論を重ねて  
ブラッシュアップ



例

ある学区内における教育の質に関する意見を親に尋ねる場合・・・

- 学校に関する意見を分類するようなカテゴリの作成・・・「学校」「先生」「課外活動」など
- 「肯定的な意見」と「否定的な意見」を分類するようなカテゴリの作成・・・「安心」「帰りが遅い」「心配」など

- テキストマイニングでは、一連の回答に含まれている情報を要約するだけで十分ということもあり得ます
- この場合、分析から得られたカテゴリが最終的なアウトプットとなります

- 本プロジェクトでは、主要な統計解析ソフトウェアのひとつである「**IBM SPSS**」の統計解析機能、およびテキストマイニング機能を用いて、公共施設関連データおよびアンケート結果の統計解析・テキスト分析を行いました。

### SPSSとは

- 顧客購買データ・医療データなど、さまざまなデータをより簡単により深く分析するための統計解析ソフトウェアで、全世界で28万人、国内で3万人が利用しており、**世界的に主要な統計解析ソフトの一つとして広く認知**されています。
- 目的・用途に合わせて用意された様々な統計解析・テキストマイニング機能が用意されており、それを用いることで、データを計算したり、データ間の関連を調べたり、様々な形式のグラフ・表へ出力するなど、様々な分析に用いることが可能です

#### 【ソフトウェアの一例】

##### **IBM SPSS Statistics**

統計解析のスタンダードソフトウェア。分かりやすいインターフェースと多種多様な統計解析手法、グラフやレポートを搭載。

##### **SPSS Text Analytics for Surveys**

自然言語処理 (NLP) テクノロジーを使用し、アンケート調査等のテキストデータの分析を行うテキストマイニングツールです。調査回答を分類し、その結果を他の調査データと統合してより優れたインサイトと統計分析が可能です

##### **IBM SPSS Modeler**

ビッグデータ、IoT 時代のデータ分析を支えるデータマイニング・ワークベンチ。構造化、非構造化データを含めあらゆるデータを統合して分析が可能。予測、パターン発見、セグメンテーションなどの分析手法を搭載。

##### **SPSS Modeler Text Analytics**

自然言語処理 (NLP) テクノロジーを使用し、テキストデータの分類/データ統合/分析を行えるテキストマイニングツールです。キーワードや同義語の辞書・ライブラリー・テンプレートなど、一連の言語リソースを搭載。



# 4.SPSSを用いた統計解析・テキストマイニング 自由記述欄の分析プロセス (1/2)



- 「①カテゴリの作成」・「②カテゴリ間の関係性の発見」のプロセスで、実際に行う分析作業は以下のとおり。

プロセス	分析の流れ	分析の位置付け	概要	分析イメージ																																													
①カテゴリの作成	アンケートの読み込み	自由記述欄の <b>大まかな内容</b> を確認し、仮説を構築	<ul style="list-style-type: none"> <li>分析対象となるアンケートの自由記述欄に記載されている内容に目を通し、意見の大まかな内容を掴み、どのような項目で整理するか仮説を立てる</li> </ul>																																														
	頻度分析	自由記述内の <b>頻出語句(出現回数)</b> を確認	<ul style="list-style-type: none"> <li>テキスト情報を自然言語解析によって「語句」に分解し、各語句の出現回数を分析することで、重要語句を抽出したり、アンケート全体の特徴や傾向を把握する</li> </ul>	<table border="1"> <thead> <tr> <th>年齢区分</th> <th>市内</th> <th>市外</th> </tr> </thead> <tbody> <tr> <td>語句数</td> <td>91</td> <td>233</td> </tr> <tr> <td>1</td> <td>雑多 (9)</td> <td>街 (26)</td> </tr> <tr> <td>2</td> <td>街 (9)</td> <td>若者 (17)</td> </tr> <tr> <td>3</td> <td>交流 (6)</td> <td>活気がある (14)</td> </tr> <tr> <td>4</td> <td>人々 (5)</td> <td>雑多 (12)</td> </tr> <tr> <td>5</td> <td>文化 (3)</td> <td>最先端 (12)</td> </tr> </tbody> </table>	年齢区分	市内	市外	語句数	91	233	1	雑多 (9)	街 (26)	2	街 (9)	若者 (17)	3	交流 (6)	活気がある (14)	4	人々 (5)	雑多 (12)	5	文化 (3)	最先端 (12)																								
	年齢区分	市内	市外																																														
語句数	91	233																																															
1	雑多 (9)	街 (26)																																															
2	街 (9)	若者 (17)																																															
3	交流 (6)	活気がある (14)																																															
4	人々 (5)	雑多 (12)																																															
5	文化 (3)	最先端 (12)																																															
テキストリンク分析	自由記述内の <b>語句間の関係性・意味のかたまり</b> を確認	<ul style="list-style-type: none"> <li>テキスト情報に含まれる語句間の関係性を視覚化し、頻度分析で把握した主要な語句を中心に、テキスト情報にどのような意味のかたまり・傾向・特徴が含まれているかを検討する</li> <li>また、似た意味を持つ語句のかたまりは「カテゴリ」として定義し、相関分析やコレスポネンス分析のインプットとして用いる</li> </ul>																																															
②カテゴリ間の関係性の発見	相関分析	<b>カテゴリ間の関係の強さ</b> を確認	<ul style="list-style-type: none"> <li>テキストリンク分析等を通じて作成したすべてのカテゴリ間の相関係数を算出することで、主要なカテゴリ同士の関係性の確認やカテゴリ間の想定外の相関関係の有無をチェックする</li> </ul>	<table border="1"> <thead> <tr> <th></th> <th>にぎやか</th> <th>まちづくり</th> <th>イベント</th> <th>イメージ</th> </tr> </thead> <tbody> <tr> <td>にぎやか</td> <td>1</td> <td>.088**</td> <td>.085**</td> <td>.006</td> </tr> <tr> <td>まちづくり</td> <td>-.088**</td> <td>1</td> <td>-.073**</td> <td>-.011</td> </tr> <tr> <td>イベント</td> <td>.085**</td> <td>-.073**</td> <td>1</td> <td>-.006</td> </tr> <tr> <td>イメージ</td> <td>.006</td> <td>-.011</td> <td>-.006</td> <td>1</td> </tr> <tr> <td>シンボル</td> <td>-.006</td> <td>-.010</td> <td>.001</td> <td>-.022</td> </tr> <tr> <td>交通</td> <td>-.024</td> <td>-.105**</td> <td>-.003</td> <td>-.014</td> </tr> <tr> <td>住民</td> <td>-.028</td> <td>.138**</td> <td>.031</td> <td>-.020</td> </tr> <tr> <td>便利</td> <td>.002</td> <td>.024</td> <td>-.010</td> <td>-.019</td> </tr> </tbody> </table>		にぎやか	まちづくり	イベント	イメージ	にぎやか	1	.088**	.085**	.006	まちづくり	-.088**	1	-.073**	-.011	イベント	.085**	-.073**	1	-.006	イメージ	.006	-.011	-.006	1	シンボル	-.006	-.010	.001	-.022	交通	-.024	-.105**	-.003	-.014	住民	-.028	.138**	.031	-.020	便利	.002	.024	-.010	-.019
		にぎやか	まちづくり	イベント	イメージ																																												
にぎやか	1	.088**	.085**	.006																																													
まちづくり	-.088**	1	-.073**	-.011																																													
イベント	.085**	-.073**	1	-.006																																													
イメージ	.006	-.011	-.006	1																																													
シンボル	-.006	-.010	.001	-.022																																													
交通	-.024	-.105**	-.003	-.014																																													
住民	-.028	.138**	.031	-.020																																													
便利	.002	.024	-.010	-.019																																													
コレスポネンス分析	<b>カテゴリと属性の関係</b> を確認	<ul style="list-style-type: none"> <li>アンケートの集計結果を散布図で表現することで、「どのカテゴリ同士が近いイメージで認識されているか」や「どの属性において特徴的に述べられているか」を視覚的に把握する</li> </ul>																																															

# 4.SPSSを用いた統計解析・テキストマイニング 自由記述欄の分析プロセス (2/2)



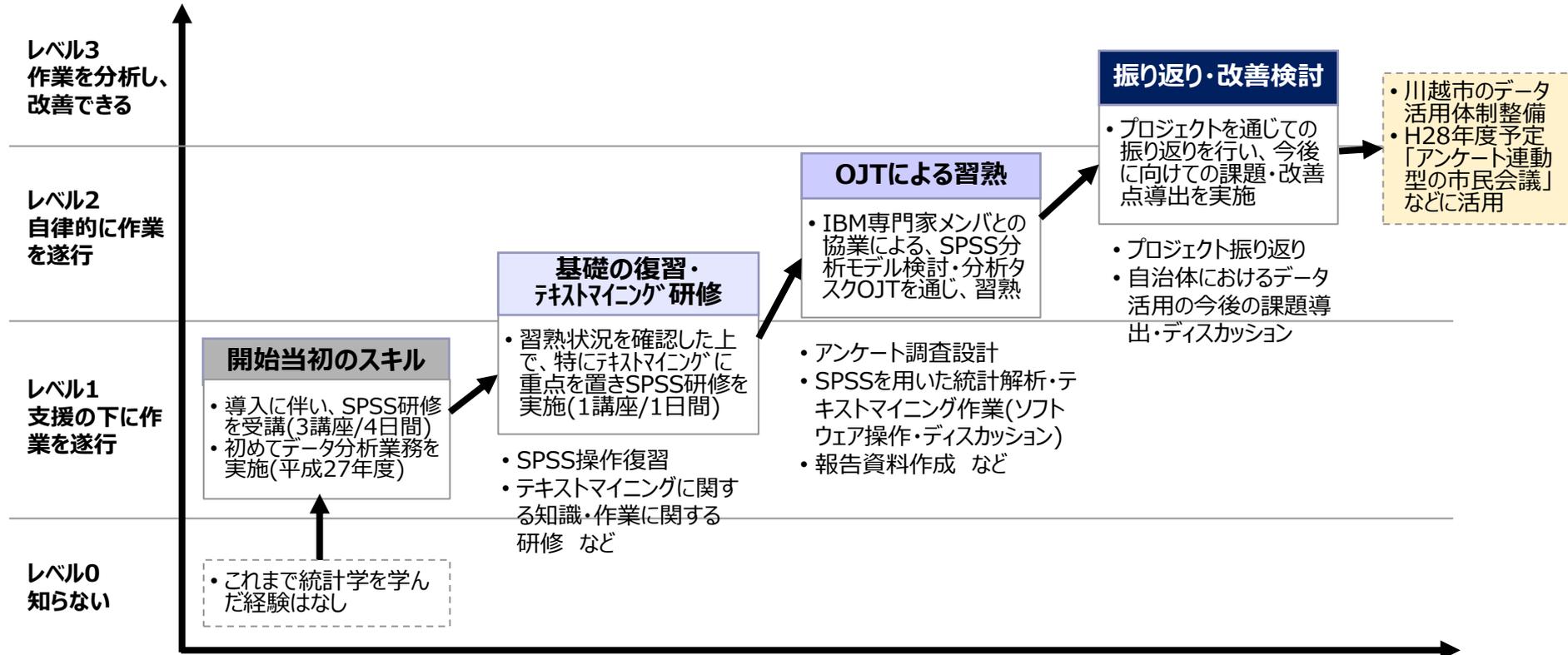
- SPSSツール群と、分析作業の対応は以下のとおり。
- 本プロジェクトでは、データアナリストが必要に応じたツールを活用し、分析を一通り実施しました。

プロセス	分析の流れ	分析の位置付け	概要	分析イメージ																																																		
①カテゴリの作成	アンケートの読み込み	自由記述欄の <b>大まかな内容</b> を確認し、仮説を構築	分析対象となるアンケートの自由記述欄に記載されている内容に目を通し、意見の傾向を把握し、仮説を立てる <b>分析ツールは必要なし</b>																																																			
	頻度分析	自由記述内の <b>頻出語句(出現回数)</b> を確認	テキスト情報を自然言語処理で分解し、各語句の出現回数を分析することで、重要語句を抽出したり、アンケート全体の特徴や傾向を把握する <b>SPSS Statistics SPSS TAFs</b>	<table border="1"> <thead> <tr> <th>年齢区分</th> <th>市内</th> <th>市外</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>雑多 (9)</td> <td>街 (26)</td> </tr> <tr> <td>2</td> <td>街 (9)</td> <td>若者 (17)</td> </tr> <tr> <td>3</td> <td>交流 (6)</td> <td>活気がある (14)</td> </tr> <tr> <td>4</td> <td>人々 (5)</td> <td>雑多 (12)</td> </tr> <tr> <td>5</td> <td>文化 (3)</td> <td>最先端 (12)</td> </tr> </tbody> </table>	年齢区分	市内	市外	1	雑多 (9)	街 (26)	2	街 (9)	若者 (17)	3	交流 (6)	活気がある (14)	4	人々 (5)	雑多 (12)	5	文化 (3)	最先端 (12)																																
	年齢区分	市内	市外																																																			
1	雑多 (9)	街 (26)																																																				
2	街 (9)	若者 (17)																																																				
3	交流 (6)	活気がある (14)																																																				
4	人々 (5)	雑多 (12)																																																				
5	文化 (3)	最先端 (12)																																																				
テキストリンク分析	自由記述内の <b>語句間の関係性・意味のかたまり</b> を確認	テキスト情報に含まれる語句間の関係性を視覚化し、頻度分析で把握した語句間の関係性を可視化し、傾向や意味のかたまり・傾向を把握する また、似た意味の語句を抽出し、相関分析やコレスポネンス分析のインプットとして用いる <b>SPSS Modeler Text Analytics (実施しなくともカテゴリは作成可能。)</b>																																																				
②カテゴリ間の関係性の発見	相関分析	<b>カテゴリ間の関係の強さ</b> を確認	テキストリンク分析等から抽出されたカテゴリ間の相関係数を算出することで、主要な相関関係の確認やカテゴリ間の想定外の相関関係の有無を確認する <b>SPSS Statistics SPSS TAFs</b>	<table border="1"> <thead> <tr> <th></th> <th>にぎやか</th> <th>まちづくり</th> <th>イベント</th> <th>イメージ</th> </tr> </thead> <tbody> <tr> <td>にぎやか</td> <td></td> <td>0.088**</td> <td>0.085**</td> <td>-0.085**</td> </tr> <tr> <td>まちづくり</td> <td>0.088**</td> <td></td> <td>0.073**</td> <td>-0.073**</td> </tr> <tr> <td>イベント</td> <td>-0.085**</td> <td>0.073**</td> <td></td> <td>-0.085**</td> </tr> <tr> <td>イメージ</td> <td>-0.085**</td> <td>-0.073**</td> <td>-0.085**</td> <td></td> </tr> <tr> <td>ジャンル</td> <td>-0.006</td> <td>-0.011</td> <td>-0.006</td> <td>-0.006</td> </tr> <tr> <td>交通</td> <td>-0.006</td> <td>-0.010</td> <td>-0.001</td> <td>-0.001</td> </tr> <tr> <td>住民</td> <td>-0.024</td> <td>-0.025</td> <td>-0.003</td> <td>-0.003</td> </tr> <tr> <td>住居</td> <td>-0.029</td> <td>-0.028</td> <td>-0.021</td> <td>-0.021</td> </tr> <tr> <td>便利</td> <td>0.000</td> <td>0.024</td> <td>-0.010</td> <td>-0.010</td> </tr> </tbody> </table>		にぎやか	まちづくり	イベント	イメージ	にぎやか		0.088**	0.085**	-0.085**	まちづくり	0.088**		0.073**	-0.073**	イベント	-0.085**	0.073**		-0.085**	イメージ	-0.085**	-0.073**	-0.085**		ジャンル	-0.006	-0.011	-0.006	-0.006	交通	-0.006	-0.010	-0.001	-0.001	住民	-0.024	-0.025	-0.003	-0.003	住居	-0.029	-0.028	-0.021	-0.021	便利	0.000	0.024	-0.010	-0.010
		にぎやか	まちづくり	イベント	イメージ																																																	
にぎやか		0.088**	0.085**	-0.085**																																																		
まちづくり	0.088**		0.073**	-0.073**																																																		
イベント	-0.085**	0.073**		-0.085**																																																		
イメージ	-0.085**	-0.073**	-0.085**																																																			
ジャンル	-0.006	-0.011	-0.006	-0.006																																																		
交通	-0.006	-0.010	-0.001	-0.001																																																		
住民	-0.024	-0.025	-0.003	-0.003																																																		
住居	-0.029	-0.028	-0.021	-0.021																																																		
便利	0.000	0.024	-0.010	-0.010																																																		
コレスポネンス分析	<b>カテゴリと属性の関係</b> を確認	アンケートの集計結果を数値で表現することで、「どのカテゴリ同士が近いイメージで認識されているのか」を可視化し、特徴的に述べられるカテゴリを抽出する <b>SPSS Categories (実施しなくともカテゴリは作成可能。)</b>																																																				

## 5.データアナリストへのトレーニング トレーニング内容と習熟度の変化イメージ

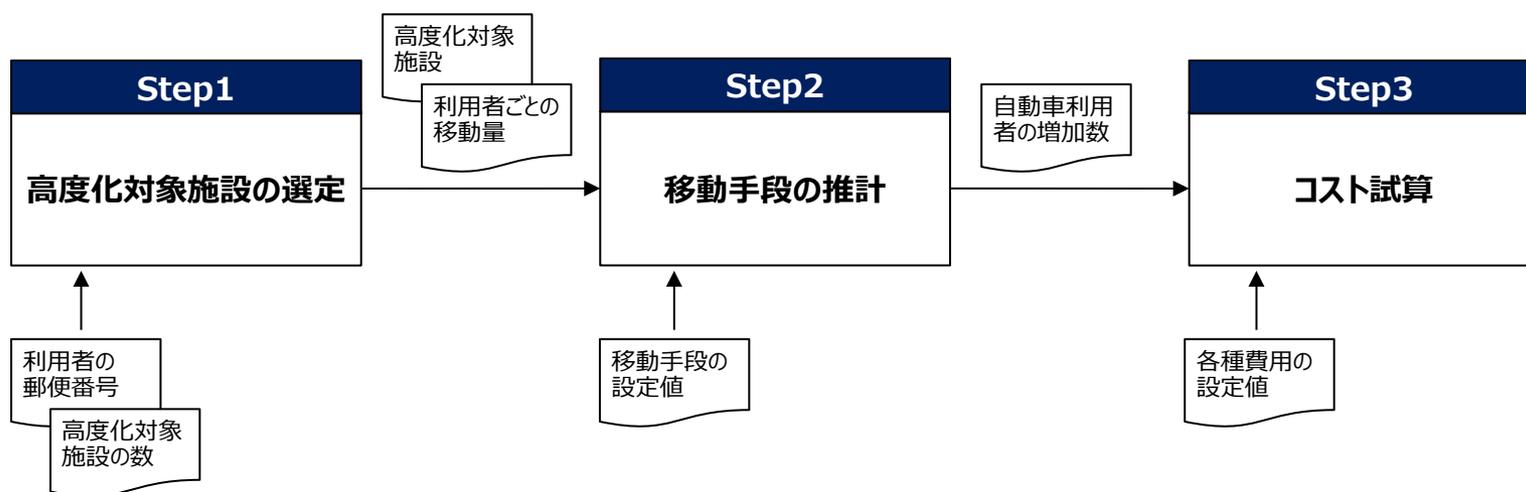
- プロジェクト開始時点において、川越市データアナリスト担当者はSPSSに関する基礎研修を修了済み(H27年度)
- そこで、まず1日掛けて集中的に基礎固めを行った後、**IBM専門家メンバとの共同分析作業を通じたOJTを中心に、データを用いた分析ノウハウや考察への習熟度向上を実施。自律的に作業が出来るようになりました。**
- また、プロジェクトの振り返りを通じ、「自治体におけるデータ活用の今後の課題」についての導出・ディスカッションを実施しました。**本プロジェクトで得た知見・ノウハウは、今後の川越市におけるデータ活用体制整備や、H28年度に予定されている「アンケート連動型の市民会議」で活かされます。**

### データアナリスト習熟度



※SPSSに関する基礎の復習・テキストマイニング研修に用いた資料については、別紙3「SPSSトレーニングセッション資料」を参照

- コスト分析は「①高度化対象施設の選定」「②移動手段の推計」「③コスト試算」の3つのステップで行いました。
- ①においては、市において高度化対象施設が決まっていないため、アンケートで取得した利用者の「郵便番号」情報を活用し、「居住地から公共施設までの距離」を指標として用いた最適化計算を行うことで、高機能化・集約化対象を仮設定し、それによるコストシミュレーションを実施しました



- 市として集約化対象施設が未決のため、利用者の居住地と公民館までの距離に基づき機械的に選定

- 利用者ごとの移動量をもとに、徒歩/自動車の閾値を設定し、自動車利用者の増加数を算出

- 各ケースごとに自動車利用者の増加数を補う施策を設定し、施設と交通施策にかかる費用を算出

### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。



- 各ステップにおける分析プロセスごとに行った主な工夫は、以下の通り

### ①高度化対象施設の選定

- 郵便番号情報を保有する利用者394件(市外・郵便番号の記載ミスを除く)についてウエイトバックを実施
- 郵便番号をもとに、利用者の居住地と公民館までの移動量を算出
- 指定した施設数を集約した時に利用者全員の移動量が最小となるように高度化対象施設を選定

### ②移動手段の推計

- 移動量に閾値を設定し、集約前後の移動手段(徒歩/自動車)の変化を算出
- 公民館の年間利用者数(約76万人)をもとにウエイトバックを行い、自動車利用者の増加数を算出

### ③コスト試算

- 集約に伴う自動車利用者の増加を補う方法として3つのケースを設定
- ケースごとに移動手段の変化を仮定し、移動手段ごとのインフラ整備費や維持・運用費を設定してコスト試算を実施

- ケース1 | 駐車場拡大型 : 自動車利用者の増加を公民館の駐車場を拡大することでカバー
- ケース2 | 公共交通活用型 : ケース1に加えて、公共交通の利用拡大でカバー
- ケース3 | 民間活用型 : ケース2に加えて、ライドシェア・駐車場シェア等の民間事業者提携でカバー

※具体的なコスト試算内容については、別紙4「コストシミュレーションツールの内容と分析結果」を参照

#### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。



- コスト分析を行う上で設定した前提条件は以下のとおり。

### 【高度化対象施設の選定における前提条件】

- 本シミュレーションでは、利用者は基本的には居住地に一番近い公民館に行くものと仮定して試算を行う
- 上記に基づき、高度化対象施設を決定する上で、制約条件には利用者の居住地と公民館の距離のみを考慮する
- 集約後の残施設数は12とする(既存17施設の $\frac{1}{3}$ ~ $\frac{1}{4}$ を削減するものとする)
- 建築後20年未満の公民館は集約対象候補施設から除外する(除外対象：南、高階、大東、伊勢原、名細)
- 削減した施設は学校等に集約して補うのではなく、他の公民館に移動することを想定する(集約化対象施設の現在利用者→他の施設に行く場合を想定)

### 【移動手段の推計における前提条件】

- 徒歩による移動時間が15分以内の場合は徒歩、15分を超える場合は自動車利用とする

### 【コスト試算における前提条件】

- コスト試算は公民館運営に関する費用(施設更新、集約化で追加整備が必要な駐車場・バス等)について分析を行うものであり、既存の駐車場・バス運営等にかかる費用は考慮していない
- 施設は築30年経過で大改修、築60年経過で建替えを行うものとする
- 2015年時点で築30年・60年を経過している施設については初年度に改修・建替え費用を計上する
- シミュレーション期間内(15年先まで)に改修・建替えが発生する施設は初年度に費用を計上する
- 施設の大改修・建替え費用は川越市公共施設マネジメント白書に則った値とする
- 集約受入施設は改修費及び維持費が1.2倍になると仮定する

#### 【本プロジェクトにおける前提】

- 本プロジェクト実施時(2015年調査時点)において、川越市として具体的に公共施設高度化・集約化検討を進めている対象施設・施設種別等は、ありません。
- 本プロジェクトの目的である「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を抽出するために検討プロセスを試行する上での、あくまで仮の検討対象として「公民館」を設定・例示したものです。

1. 本プロジェクトにおける前提
2. プロジェクトの概要と目的
  1. プロジェクト概要
  2. 目的とゴール
3. プロジェクトのアプローチ
  1. アプローチ
  2. スケジュール
  3. 体制・役割分担
4. 公共施設高度化検討におけるデータ活用検討のポイント
  1. 全体像
  2. データ活用検討のポイント
    1. 分析視点・対象施設の検討
    2. 利用者アンケート調査・設計
    3. 調査実施・データ化作業
    4. SPSSを用いた統計解析・テキストマイニング
    5. データアナリストへのトレーニング
    6. 集約対象候補選定・コストシミュレーション
5. 自治体におけるデータ活用促進に向けた考慮点・気づき
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題
6. 参考資料
  1. 別紙1：公共施設アンケート調査票
  2. 別紙2：川越市公民館利用者アンケート調査結果
  3. 別紙3：SPSSトレーニングセッション資料
  4. 別紙4：コストシミュレーションツールの内容と分析結果

# 自治体におけるデータ活用促進に向けた考慮点・気づき

## 1. 住民の生の声を行政に活かす上の考慮点・課題

- 組織の枠を超えた情報共有・連携の仕組みづくり
  - 自治体・民間企業(交通・医療他)・研究機関など
- オープンデータ化の促進
  - 個人情報管理・情報セキュリティ・過去データも含めた保有情報の活用ルールや一元管理の仕組みづくり
- 効率的にデータ活用を行うための仕組みづくり
  - データ化作業体制の整備(内製or委託)
- 自由記述情報を活用できる仕組みづくり
  - データ処理に適したハード・ソフト環境の整備(統計解析・テキストマイニング)

## 2. 生の声を分析するためのアンケート設計・分析の考慮点・課題

## 3. データアナリスト育成・活用の考慮点・課題

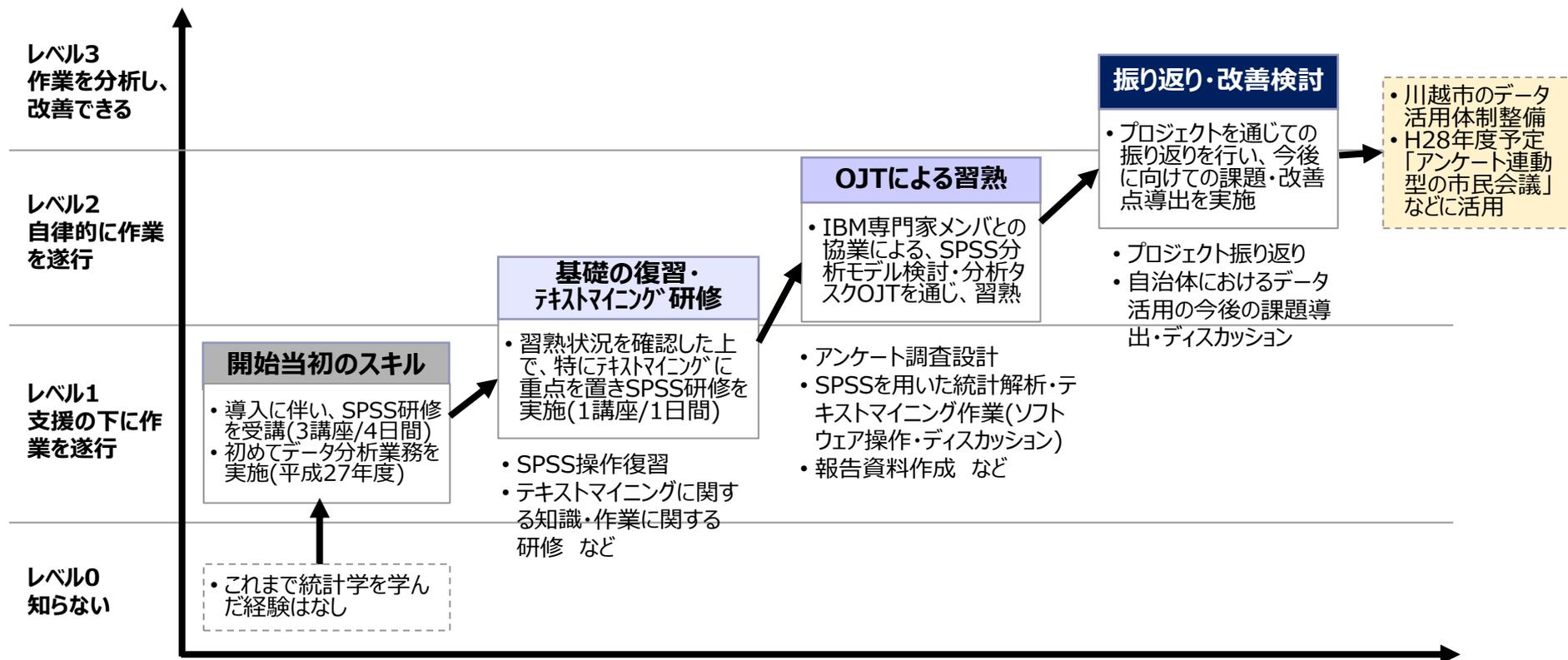
- データアナリスト複数体制の整備
  - データアナリストの異動・離職を考慮した体制
- データアナリストの育成プロセスの整備・内製化
- 分析作業に集中できる環境づくり
- 関係者に対し、データ活用の価値認識を向上させる取り組み
  - トライアル事例・クイックウィン事例の蓄積、組織内への広報活動

### 検討・作成中

- 本プロジェクトを通じて得られた・他展開が可能な「自治体におけるデータ活用やデータアナリスト育成における考慮点・課題」を最後にまとめる項。
- 検討作業および担当者の声を基に、以下の観点で整理
  1. 住民の生の声を行政に活かす上の考慮点・課題
  2. 生の声を分析する上での調査設計・分析の考慮点・課題
  3. データアナリスト育成・活用の考慮点・課題

- プロジェクト開始時点において、川越市データアナリスト担当者はSPSSに関する基礎研修を修了済み(H27年度)
- そこで、まず1日掛けて集中的に基礎固めを行った後、**IBM専門家メンバとの共同分析作業を通じたOJTを中心に、データを用いた分析ノウハウや考察への習熟度向上を実施。自律的に作業が出来るようになりました。**
- また、プロジェクトの振り返りを通じ、「自治体におけるデータ活用の今後の課題」についての導出・ディスカッションを実施しました。**本プロジェクトで得た知見・ノウハウは、今後の川越市におけるデータ活用体制整備や、H28年度に予定されている「アンケート連動型の市民会議」で活かされます。**

データアナリスト習熟度



※SPSSに関する基礎の復習・テキストマイニング研修に用いた資料については、別紙3「SPSSトレーニングセッション資料」を参照

#	項目	説明	記入欄
1	担当者のスキル・経験	<ul style="list-style-type: none"> <li>本プロジェクト以前のスキル・受けたデータアナリスト研修・経験について</li> </ul>	<ul style="list-style-type: none"> <li>これまで統計学を学んだ経験はなく、平成27年度に初めてデータ分析を担当・実施</li> <li>27年度にSPSS statisticsの導入に伴い、講習を受講（3講座、4日間受講）</li> </ul>
2	本プロジェクトを通じての担当者の感想	<ul style="list-style-type: none"> <li>本プロジェクトを通じての感想や、得られたものを記述して下さい</li> </ul>	<ul style="list-style-type: none"> <li>アンケート作成・配布・回収方法などで考慮すべき点や、SPSSを活用したテキストマイニング分析の手順・ノウハウの理解・習熟が進んだ</li> <li>短期間・小規模の活動にもかかわらず、分析を通じた気付きも多くあり、社会資本マネジメント担当者とのディスカッションも進んだ。自治体におけるデータ活用の有効性・必要性を感じた</li> <li>アンケート作成段階での仮説を立てることについて、日々収集する情報といかにデータを関連付けて考えられるかが重要と感じた</li> <li>本データ分析で学んだノウハウを活用し、市民の方の声を政策に反映するしくみとして「アンケート連動型の市民会議」の実施を検討している（1月予算内示で決定）</li> </ul>
3	今後の課題	<ul style="list-style-type: none"> <li>自治体におけるオープンデータ活用や、データアナリスト育成を今後推進していく上で課題と感じている事を記述して下さい</li> </ul>	<p><b>【組織体制】</b></p> <ul style="list-style-type: none"> <li>データアナリストの配置数が少なく、ディスカッションの相手・時間を見つけるのが難しい。</li> <li>他の兼務作業を行いながら、分析作業に集中できる環境・時間をどう作るかが課題</li> <li>個人情報やセキュリティのデータ活用についても柔軟に活用できる組織づくりが課題</li> <li>他自治体との比較、過去の実績・事例など今までの意思決定の判断基準に加えて、データに基づく客観的な判断を行う体制づくりが必要</li> </ul> <p><b>【プロセス】</b></p> <ul style="list-style-type: none"> <li>今後の増員や担当者の異動も考慮すると、データアナリスト育成プロセスの整備が急務</li> <li>テキストマイニングなどデータ活用に対する原課の認知が低い。価値を理解してもらえるよう周知・実績作りが必要</li> <li>解析に使えるような過去データは多数あるが、アンケート収集時に「他目的利用をしない」としているものが多く、活用を進める上で課題となっている</li> <li>市の情報には紙媒体のものが多く、どのようにデータ化するかは課題。今後、データ活用することを前提とした様式を策定し、利用推進していく必要がある</li> <li>庁内データを一元的に管理するしくみづくりが必要。また、官民に関わらず必要なデータを必要な職員が蓄積する意識醸成が課題</li> </ul> <p><b>【IT】</b></p> <ul style="list-style-type: none"> <li>データ処理に時間が掛かる。相応のスペックの機器が必要</li> <li>専門的なソフトを導入するためのコストが課題</li> </ul>

