



第1回 TQ E問題解決オンラインセミナー

高等学校における統計・情報教育における新たな展開



令和2年12月21日

総務省統計研究研修所

長 尾 伸 一



総務省統計研究研修所
Statistical Research
and Training Institute,
Ministry of Internal Affairs and Communications

総務省における統計教育への取組

・小中学生向け学習サイト



・高校生向け学習サイト



令和4年度から高等学校の学習指導要領が改訂されることを踏まえ、新設される「情報Ⅰ」、「情報Ⅱ」及び「地理総合」等、統計・情報に係る範囲について、**統計リテラシーの向上**及び**公的統計の利活用推進**を目的とした統計教育用教材の開発を進めているところ

総務省における統計教育への取組

・【統計学習の指導のために「先生向け」】の紹介

統計学習の指導のために **先生向け** | [統計局ホーム](#) | [統計局の紹介](#) | [ご意見・ご感想](#) |

授業モデル	補助教材	学校における統計教育の位置づけ	リンク集	サイトの歩き方	
授業モデル  新しい学習指導要領で大幅に拡充が図られた統計学習に関して、統計局が実施している統計調査の結果等を素材として用いた実践的な授業モデルを掲載しています。	補助教材  このコーナーに掲載した資料は、主に中学校や高校で先生方が「統計」について指導を行う際、その導入あるいは発展をサポートするための素材として統計局で作成したものです。	学校における統計教育の位置づけ  新学習指導要領及び学習指導要領解説から統計教育に関わりが深いと思われる部分を抜粋しましたので、ご活用ください。	リンク集  e-stat（政府統計の総合窓口）、都道府県等の統計関連サイト、諸外国統計局の学習サイト等へのリンク集です。	サイトの歩き方  サイト内のページリンクを一覧で記載しています。	ご意見・ご感想  統計学習の現場で役に立つと思われるコンテンツに関するアイデア等について、積極的にフィードバックいただければ幸いです。

<https://www.stat.go.jp/teacher/index.html>

総務省における統計教育への取組

高校生のための統計学習教材の提供(令和2年7月～)

統計学習の指導のために **先生向け** | [統計局ホーム](#) | [統計局の紹介](#) | [ご意見・ご感想](#)

授業モデル	補助教材	学校における統計教育の位置づけ	リンク集	サイトの歩き方
-------	------	-----------------	------	---------

[統計学習の指導のために](#) > [補助教材](#) > [総合学習のための補助教材](#) > 「[高校生のための統計学習教材](#)」

総合学習のための補助教材

「小学生のための 統計ってなあに」	「中学校のための統計 社会が、自然が、生活がみえる統計」	「高校生のための統計学習教材」
-----------------------------------	--	---------------------------------

「高校生のための統計学習教材」

表紙

[表紙 \(PDF: 143KB\)](#)

第1講 1. 記述統計 2. 標本調査の考え方 3. 推測統計 4. ビッグデータの活用と分析

[第1講 \(PDF: 4,517KB\)](#)

第2講 多変量データの扱い

[第2講 \(PDF: 2,138KB\)](#)

第3講 1. データから見る地域特性 2. GISの活用

[第3講-1 \(PDF: 1,322KB\)](#) [第3講-2\(1\)地図やGISで捉える地域 \(PDF: 1,928KB\)](#) [第3講-2\(2\)地域課題の分析事例 \(PDF: 5,595KB\)](#)

第4講 1. データから見る社会・経済情勢 2. 課題解決のためのデータ分析

[第4講-1 \(PDF: 5,626KB\)](#) [第4講-2 \(PDF: 2,099KB\)](#)

<https://www.stat.go.jp/teacher/comp-learn-03.html>

今年度提供する教材のポイント

■令和4年度からの新学習指導要領に準拠、本年は「情報Ⅱ」の範囲を中心に作成

■公的統計等「実際のデータ」を活用

■新設する「情報Ⅰ」「情報Ⅱ」に対応し、プログラミングを重視
R, Pythonが記述できるよう工夫

■「データ」及び「プログラムのコード」を提供することで誰でも推計結果を再現可能

■高校生が興味を持てるような題材も活用

教材の事例 「世界の統計」から



The screenshot shows the official website of the Japanese Statistical Agency (総務省統計局). The header includes a Google Custom Search bar and navigation links such as '採用情報', 'リンク集', 'ご意見・お問合せ', 'サイトマップ', '文字サイズ等の変更', and 'English'. A main navigation bar contains links to 'ホーム', '実施中の調査', '統計データ', 'よくある質問', '統計研究研修', '広報・募集', and '組織紹介'. Below this, a breadcrumb trail reads 'ホーム > 統計データ > 世界の統計'.

On the left sidebar, under '総合統計書のご案内', there is a link to '本書の内容' (Contents of this book) and a list of items: '利用上の注意' (Notes on use), '目次' (Table of contents), 'バックナンバー' (Back numbers), and '正誤情報' (Correction information). Below this are links for 'Q&A' and '問合せ先' (Contact information).

The main content area is titled '世界の統計' (World Statistics). It features a section for '世界の統計2020' (World Statistics 2020) with the following text:

「世界の統計」は、国際社会の実情や世界における我が国の位置付けを、統計データを用いて正確にみることを目的に編集したものです。

本書は、国際比較の観点から国際機関の提供している統計データを出典資料として、世界各国の人口、経済、社会、環境といった分野のデータを抽出し、約140の統計表にまとめました。

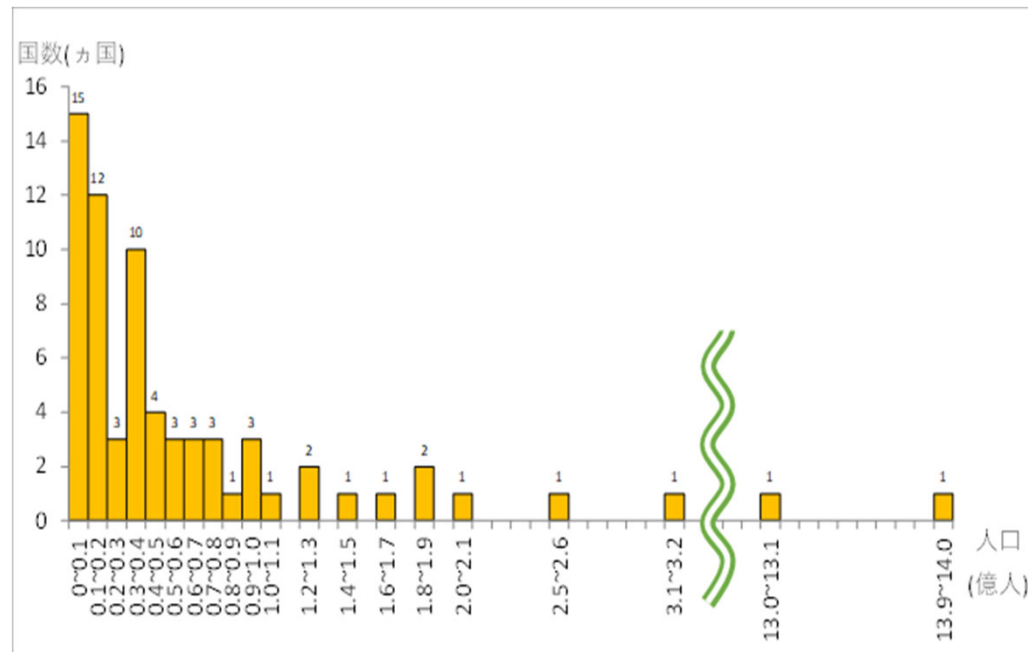
To the right of the text is a book cover for '世界の統計 2020'. The cover features a photograph of a coastal town with a rainbow in the sky. Below the book cover, the following text is provided:

表紙写真：プーリア州ポリニャーノ・ア・マーレ（イタリア）、「ポリニャーノ・ア・マーレで美しい景色を満喫」
(c) Azuma Hotta

<https://www.stat.go.jp/data/sekai/index.html>

「世界の統計」からの分析事例

<人口>

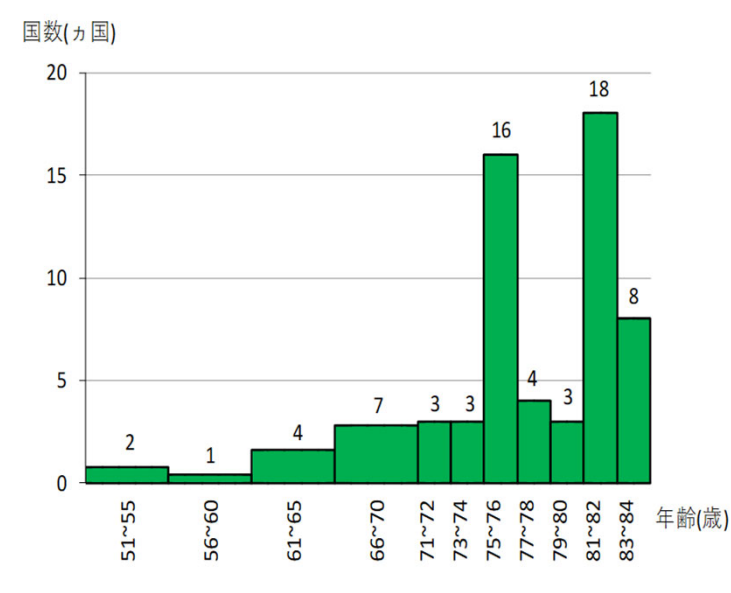
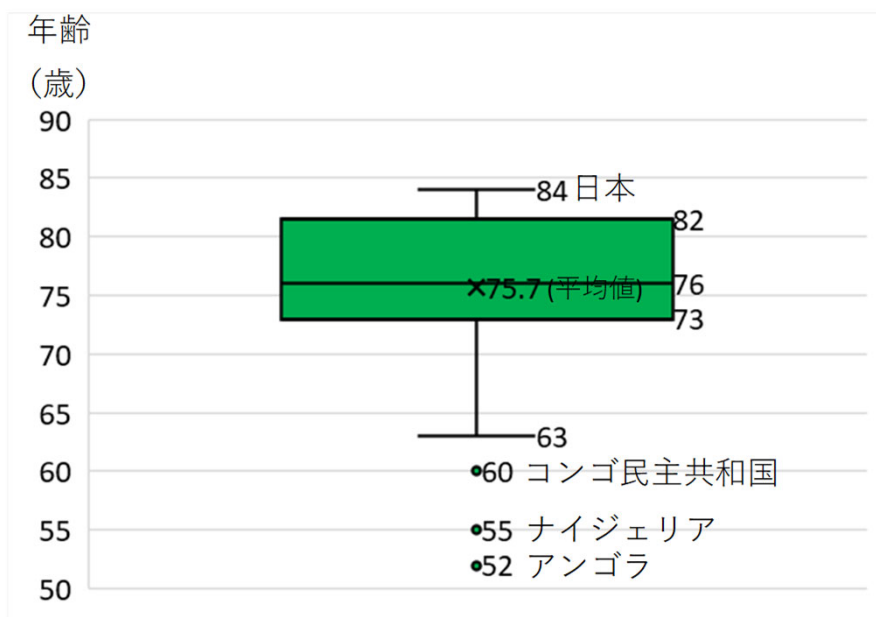


基本統計量 (億人)	
平均	0.92
標準誤差	0.28
中央値 (メジアン)	0.35
標準偏差	2.28
分散	5.22
尖度	26.91
歪度	5.11
範囲	13.97
最小	0.0033
最大	13.97
合計	63.58
データの個数	69

世界の統計から収集できる69か国の人口のヒストグラム。1000万人を下回る国が最も多い（15か国）ことが分かる。1億人を超える国は12か国あり、中国が13億97百万人と最も多く、次いでインドとなっており、両国ともに10億人を超えて、他の国に比べ圧倒的な数となっている。ちなみに、日本は2018年時点で、約1億27百万人と世界の中で10番目に人口が多い国となっている。

「世界の統計」からの分析事例

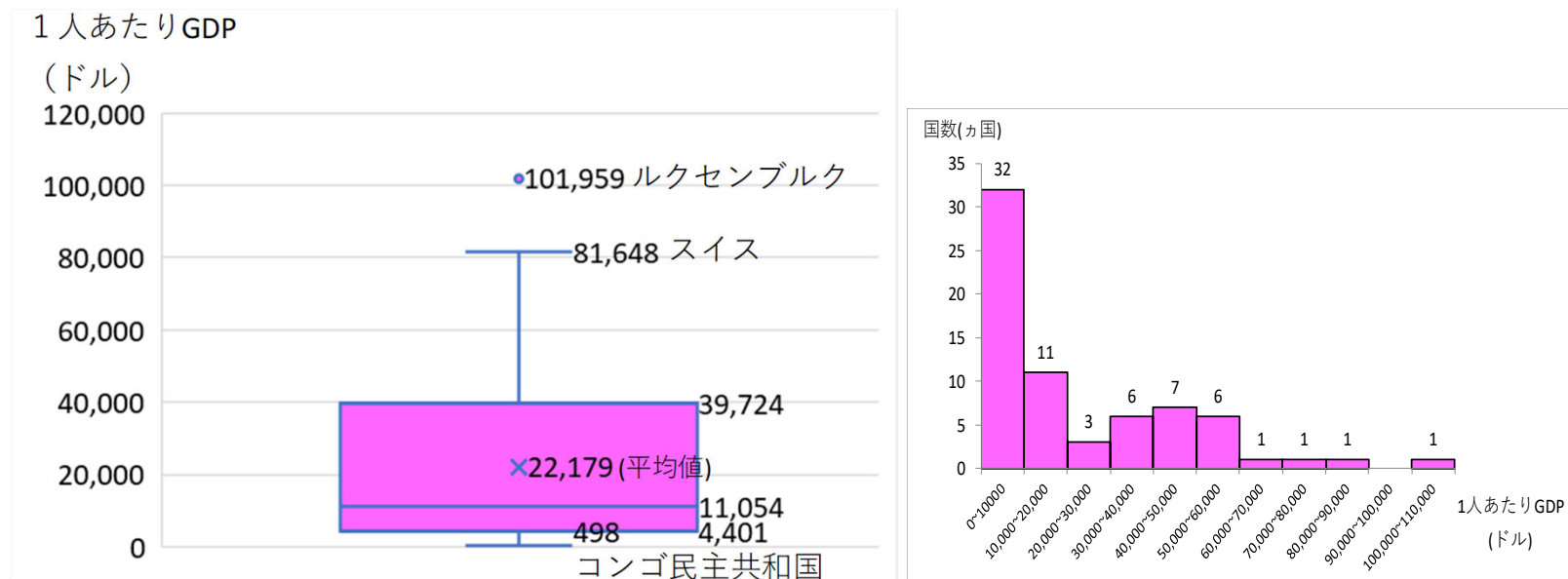
<平均余命>



平均余命は、平均値が75.7歳となっており、山が2つに分かれていることが特徴。属性の異なる集団が混在している場合に、このような形状となることが知られている。データの分布は右に偏った形状となっている。

「世界の統計」からの分析事例

<1人あたりGDP>

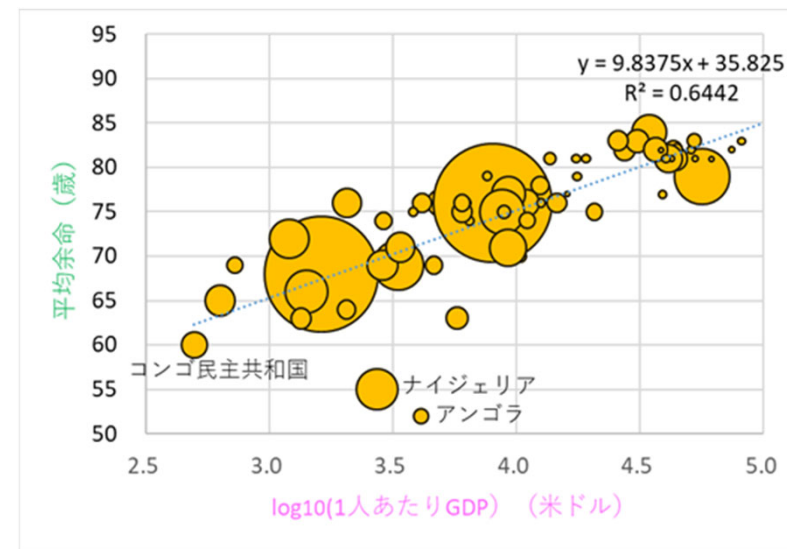
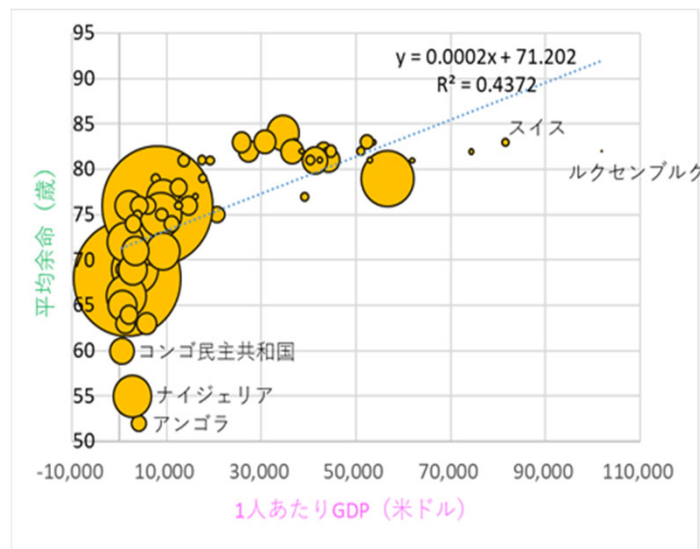


1人あたりGDPは、人口と同様に右裾の長い分布となっていることが分かる。多くの国で1人あたりGDPの値が低い一方で、少数の国で1人あたりGDPの値が高くなっている。所得や貯蓄などお金に関するデータはこのように右裾の長い分布となることが多い。

「世界の統計」からの分析事例

<人口, 1人あたりGDP, 平均余命>

一人あたりGDPと平均余命の散布図



1人あたりGDPと平均余命を縦軸と横軸にプロットした散布図である。左の図は、1人あたりGDPと平均余命をプロットしたもので、1人あたりGDPが高くなるほど、平均余命が高くなるという傾向をみることができる。右の図は横軸の1人あたりGDPを対数にとったグラフとなっている。このように対数変換すると両者が線形関係に近くなり、相関係数が高くなる。

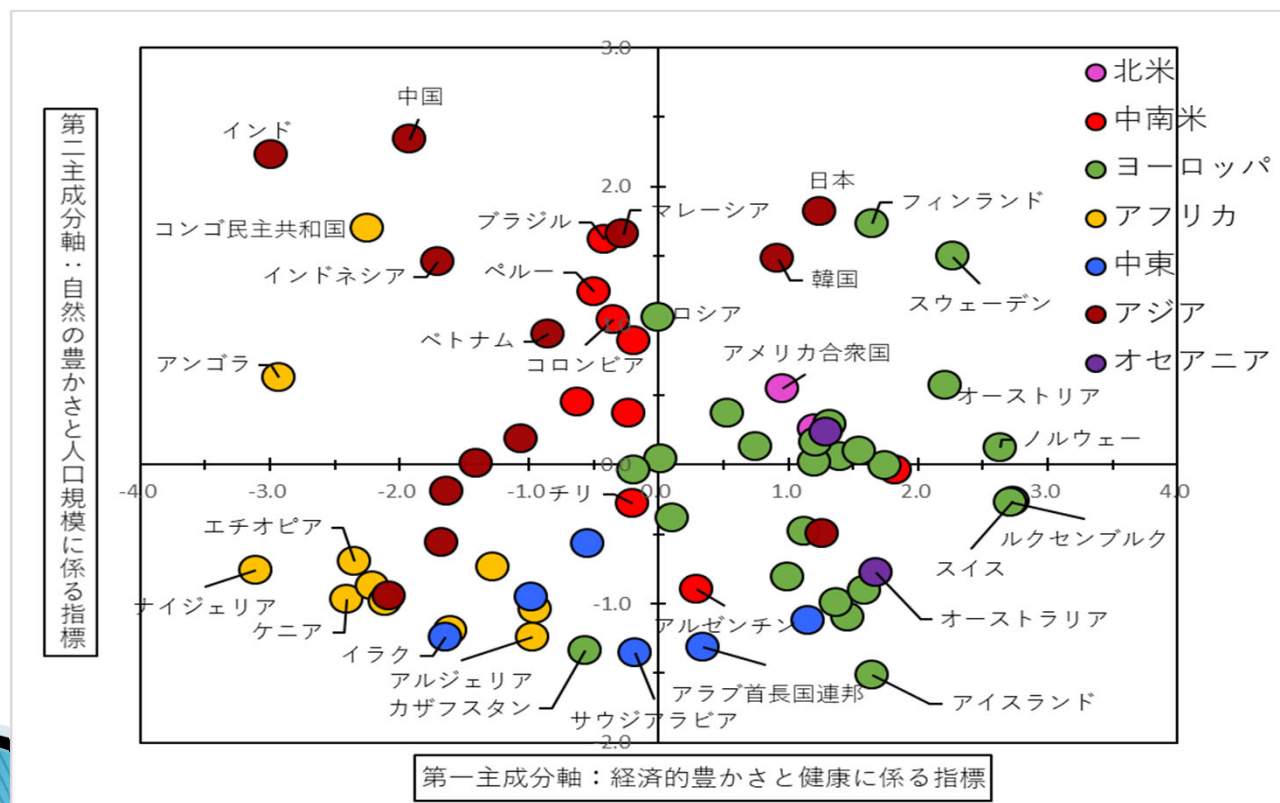
<https://www.stat.go.jp/teacher/dl/pdf/c3learn/materials/third/dai4-1.pdf>

「世界の統計」からの分析事例

<5変数の主成分分析>

「人口」(対数), 「平均余命」, 「1人当たりGDP」(対数), 「単位人口人あたり医師数」, 「森林率」の5変数で主成分分析, 推計された主成分得点を二次元にプロット。

第一主成分軸は、「1人当たりGDP」, 「平均余命」, 「人口1000人あたり医師数」が寄与。第二主成分軸は、「森林率」と「人口」が寄与。



「統計today No.159」の紹介



統計Today No.159

高校生のための統計学習教材の提供 ～プログラミング教育の開始とデータ分析力の向上に資することを目指して～

総務省統計研究研修所統括教授 長尾 伸一
(共同執筆) 同研究所統計技術向上支援課技術支援第一係 中山 拓人

総務省では、初等教育から高等教育、社会人教育までの各段階における統計リテラシーの向上に向けて、無料学習サイト・教材等の開発、小・中・高校等の段階別の教員向けコンテンツ等の提供を行ってきました。

この度、令和4年度から高等学校の学習指導要領が改訂されることを踏まえ、新設される「情報I」、「情報II」などに含まれるプログラミング、情報活用能力の向上や、公的統計を活用したデータ分析力の向上を目的として、統計学習教材を作成しました。

<https://www.stat.go.jp/teacher/comp-learn-03.html>

近年、Hans Rosling氏らが執筆した「FACTFULNESS¹」が日本語訳され話題になっていますが、データに基づき社会を読み解くことの重要性がますます広く認識されるようになってきました。今回提供する教材は、公的統計を始めとした「実際のデータ」を利用し、数理的な知識の習得ばかりでなく、データに付随する多様な知見を伝えることで、データ分析力を向上させることを目的としています。文部科学省が公表した新たな学習指導要領の内容を踏まえ、統計に係る重要な内容をできるだけ網羅するよう工夫しているほか、必修化されるプログラミングの理解を促進するためRやPythonのコードについても、可能な範囲で提供しています。

¹ Hans Rosling, Ola Rosling, Anna Rosling Rönnlund, (訳) 上杉周作, 関美和 2019.1 日経BP社

今回は、この教材に掲載している統計学習の具体的な内容について、事例を紹介します。ここで紹介するデータは総務省統計局が提供している「世界の統計2019」等²にも掲載されている諸外国の統計データを用いて、5種類の国別データ「人口」、「1人当たりGDP」、「平均余命³」、「人口1000人当たり医師数」、「森林率（陸地に占める森林面積割合）」からヒストグラム、箱ひげ図、散布図、主成分分析の結果を作成しました。

² 今回の分析に当たっては、「世界の統計」のほか、以下からも収集した（データは表1（エクセル：17KB））。

人口：UN, World Population Prospects, 1人当たりGDP：UN, National Accounts Main Aggregates Database,
平均余命：WHO, Global Health Observatory, 単位人口当たり医師数：The World Bank, World Development Indicators
森林率：FAO, Global Forest Resources Assessment

³ ここでの平均余命は0歳児の平均余命のことで、0歳児の平均余命のことを平均寿命ともいう。

<https://www.stat.go.jp/info/today/159.html>

テキストマイニングの手法を用いた事例

▶ テキストマイニングとは

テキストマイニングとは、大量のテキストデータから重要な情報を組み上げる技術のこと

自然言語解析の手法を使って、文章を単語に分割し、それらの出現頻度や相関関係といった統計的手法を用いて分析することで有益な情報を抽出

ビッグデータの活用において、「非構造化データ」とよばれる従来は取り扱いが困難であった情報を解析できる

テキストマイニングの手法を用いた事例

▶ 形態素解析とは

文章を構成する単語を最小単位に分解し、文章を分割する作業のことを指します。

<例> 私は大学で統計の勉強をします

私	は	大学	で	統計	の	勉強	を	し	ます
名詞	助詞	名詞	助詞	名詞	助詞	名詞	助詞	動詞	助動詞

形態素に分割したあとは必ずいずれかの品詞に分類される
形態素はそれ以上細分化してしまうと意味を成すことのできない最小単位

テキストマイニングの手法を用いた事例

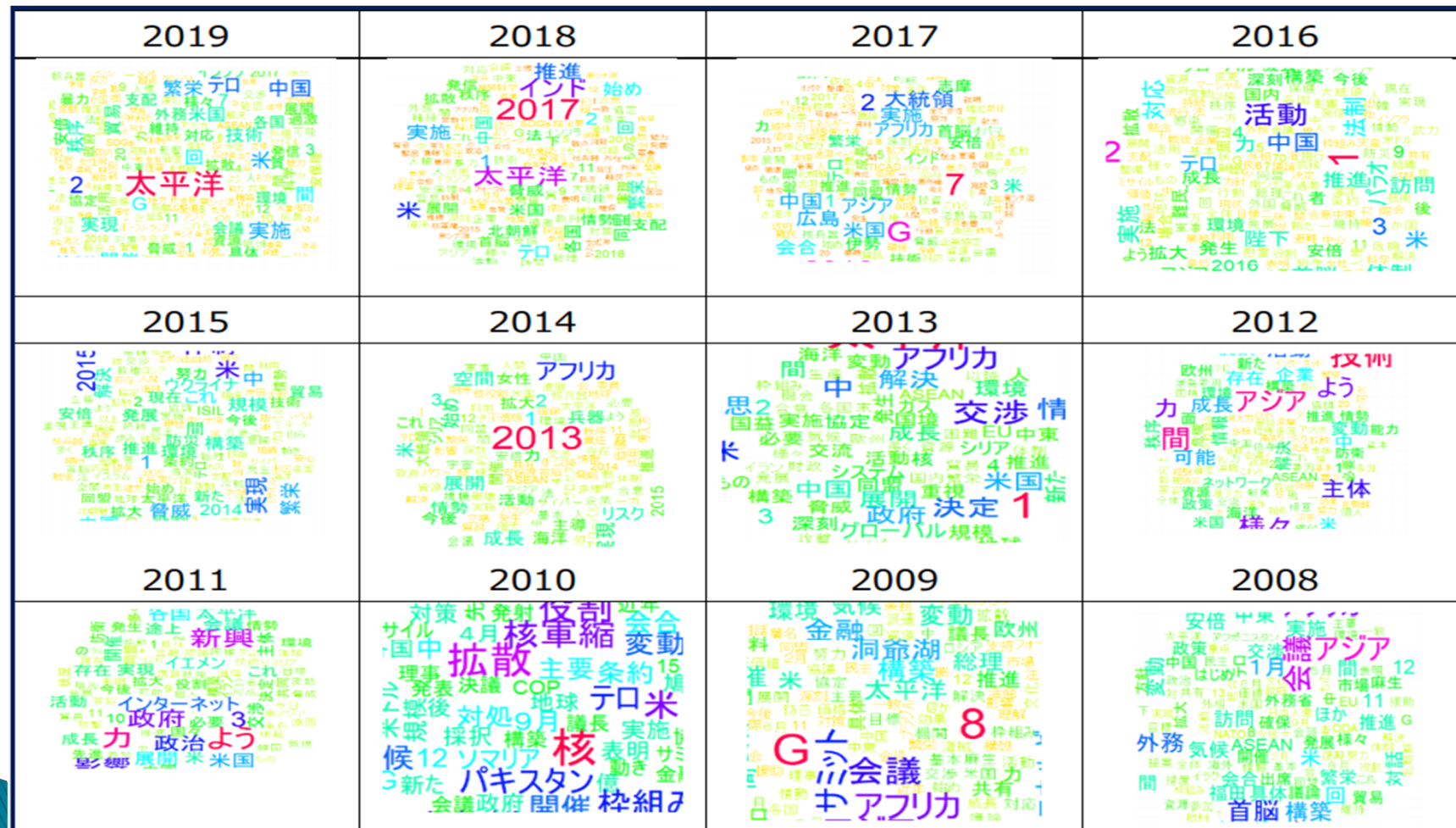
▶ 単語の出現頻度から分析する主な手法

出現頻度グラフ・ワードクラウド
単語の出現頻度を集計(グラフ化)

共起ネットワーク
同時に出現しやすい単語同士をネットワークでつないで可視化

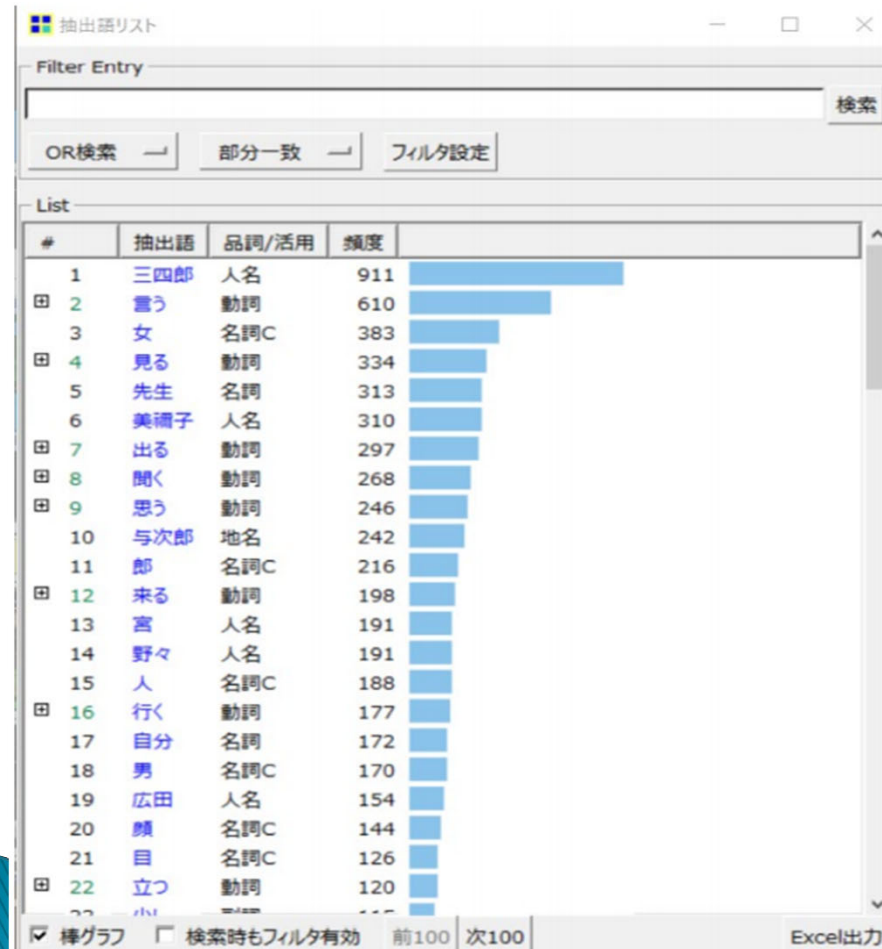
対応分析(コレスポンデンス分析)
属性情報と出現単語との対応関係を同じ座標にマッピングしその位置関係から属性の傾向を把握

テキストマイニングの手法を用いた事例 ワードクラウド(外交青書から)



テキストマイニングの手法を用いた事例 頻出単語グラフ(夏目漱石の小説から)

「三四郎」の頻出単語

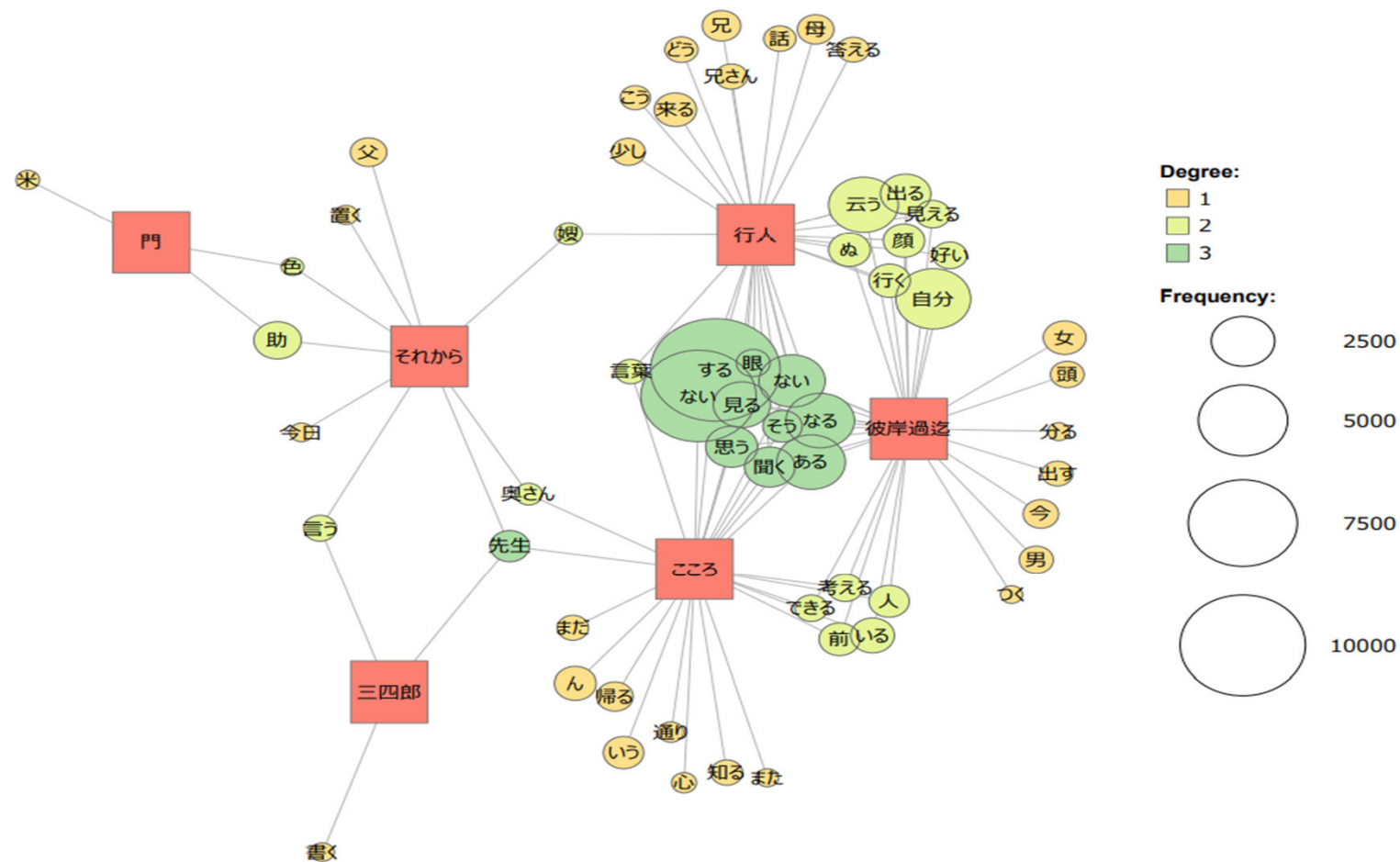


「こころ」の頻出単語



<https://www.stat.go.jp/teacher/dl/pdf/c3learn/materials/third/dai1.pdf>

テキストマイニングの手法を用いた事例 共起ネットワーク(夏目漱石の小説から)



テキストマイニングの手法

新たな取り組み 対応分析の事例(有名アーティストの歌詞を分析)

- ▶ APIでデータ収集
UTA-NETサイトから歌詞データをダウンロード
(著作権に留意)
- ▶ 形態素解析
形態素解析エンジンとしてMeCabを利用
- ▶ 頻出単語のグラフ化
- ▶ 対応分析(コレスポネンス分析)

※PCの計算資源
を大量に消費

テキストマイニングの手法を用いた取組 対応分析(ドラマ主題歌担当アーティストから)

平成22年度～直近までのNHK朝の連続ドラマ主題歌を担当したアーティスト19組の歌詞を分析

	前期			後期		
	タイトル	主題歌	アーティスト	タイトル	主題歌	アーティスト
令和2年度	エール	星影のエール	GReeeeN	おちょやん	泣き笑いのエピソード	秦基博
令和元年度	なつぞら	優しいあの子	スピッツ	スカーレット	フレア	Superfly
平成30年度	半分、青い。	アイデア	星野源	まんぷく	あなたとトゥラッタッタ♪	DREAMS COME TRUE
平成29年度	ひよっこ	若い広場	桑田佳祐	わろてんか	明日はどこから	松たか子
平成28年度	とと姉ちゃん	花束を君に	宇多田ヒカル	べっぴんさん	ヒカリノアトリエ	Mr.Children
平成27年度	まれ	オリジナル曲	澤野弘之	あさが来た	365日の紙飛行機	AKB48
平成26年度	花子とアン	にじいろ	絢香	マッサン	麦の唄	中島みゆき
平成25年度	あまちゃん	オリジナル曲	大友良英	ごちそうさん	雨のち晴レルヤ	ゆず
平成24年度	梅ちゃん先生	さかさまの空	SMAP	純と愛	いちばん近くに	HY
平成23年度	おひさま	おひさま	平原綾香	カーネーション	カーネーション	椎名林檎
平成22年度	ゲゲゲの女房	ありがとう	いきものがかり	てっぱん		葉加瀬太郎

以下の分析は、該当アーティストが発表した全ての楽曲を対象としている。おおよそ、1組当たり200曲～

テキストマイニングの手法を用いた事例

中島みゆきさんの頻出単語



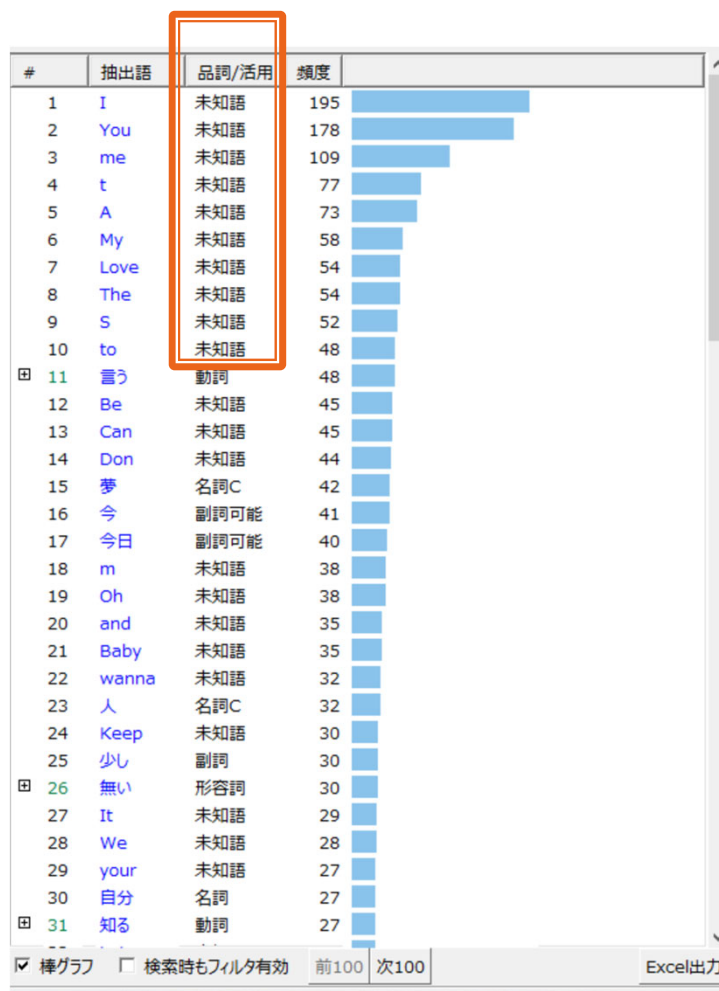
テキストマイニングの手法を用いた事例

星野源さんの頻出単語



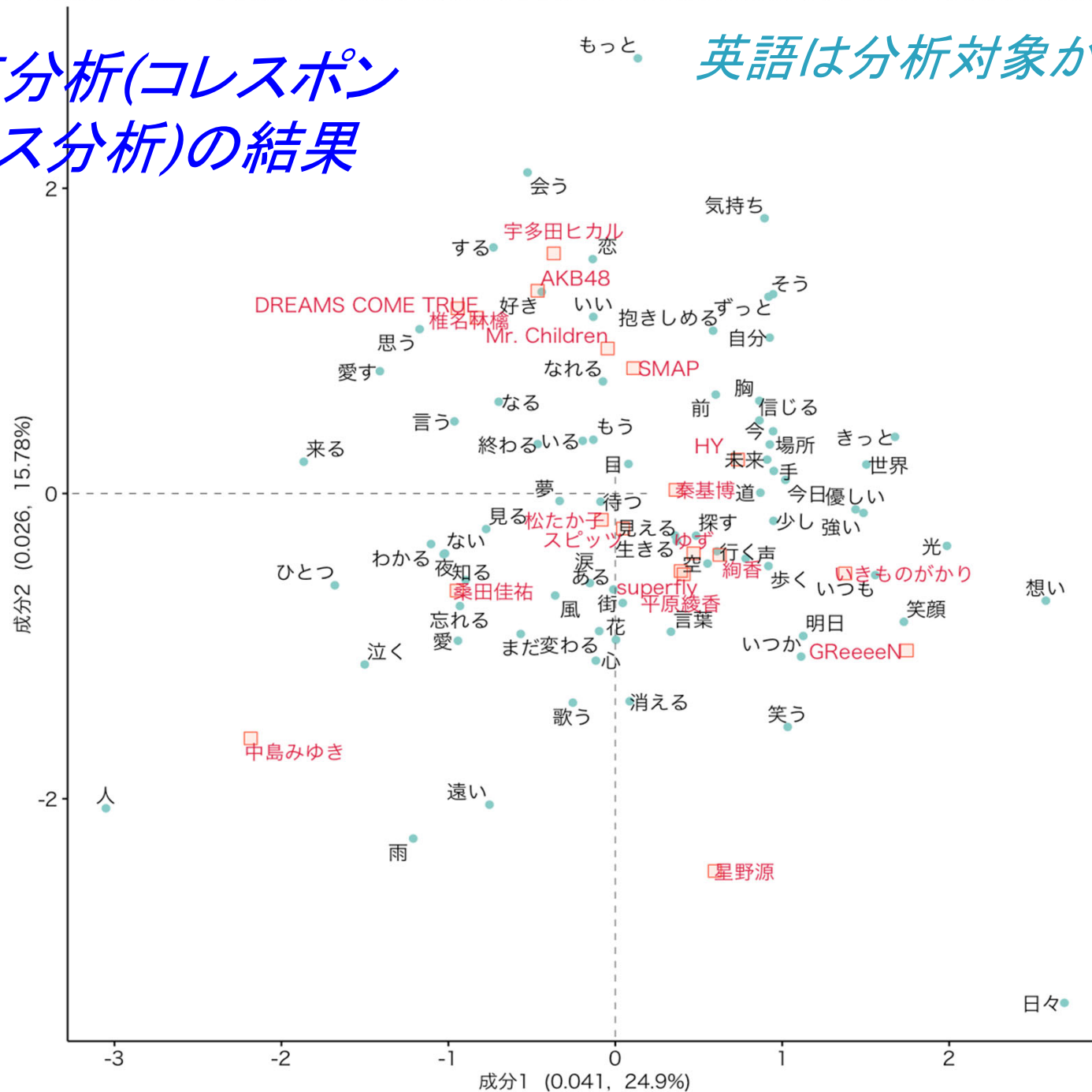
テキストマイニングの手法を用いた事例

宇多田ヒカルさんの頻出単語

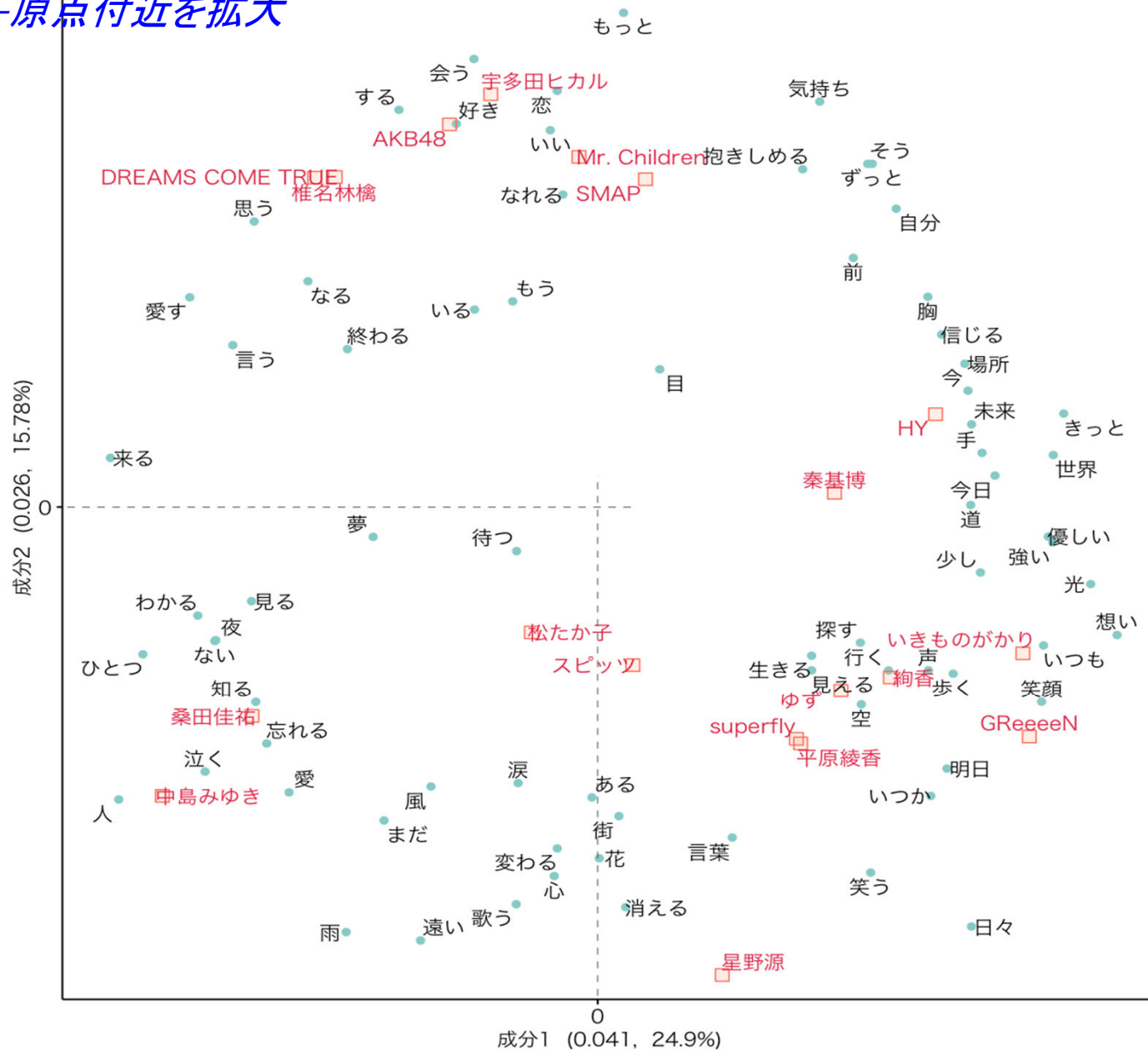


対応分析(コレスポ デンス分析)の結果

英語は分析対象から除外



対応分析(コレスポンデンス分析) の結果-原点付近を拡大



今回の事例研究からの得られた課題

- ▶ 形態素解析の結果から得られた単語の品詞のうち、どこまでを分析に利用するか。
⇒ 具体的には、「英語」の文章をどう扱うか
- ▶ 分析対象とする集団の属性は適当か。
⇒ 具体的には、「歌詞を作詞・作曲するアーティスト」と「専門家から楽曲を提供されているアーティスト」を一緒にしたことは、特性の分析を攪乱させる要因となっていないか

⇒ 分析の目的を明確化することが重要

実際のデータを利用することで見えてくる課題もある。

目的を明確にし、それを解決するためにはどうするかを考え、新たな課題が見つかった場合に、さらに検討を深める。

教材について

- ▶ データサイエンス教育にたくさんの実績を持つ(株)Rejouiに研究委託

- ▶ 各種統計分析手法の紹介

⇒文部科学省が作成した高等学校情報科「情報Ⅱ」教員研修用教材を参考にして作成

テキストマイニングのほか、

- ▶ 画像解析

- ▶ 音声解析

- ▶ データの可視化

高校生が関心を持ちそうな題材を提供

高等学校のPC及びネットワーク環境について

- ▶ PCの性能、OS(windows,Mac,etc.)に依存しない教材
- ▶ 各学校ごとに異なるネットワーク環境に依存しない教材
- ▶ ソフトウェアのインストールの可否
⇒ Python, R, その他のプログラミング言語の利用を
想定するが、クラウドでの利用も含めて検討

様々な可能性を考慮して検討しているところ

数学の予備知識について

- ▶ 大学で学習する「線形代数」、「多変数の微分・積分」についての理解を前提としない

⇒「数学ができないから統計は理解できない」という思い込みを払拭したい

⇒ただし、関心のある人向けにできるだけ簡単に数学的な説明を付録として加えたい

- ▶ 可視化を重視し、グラフや地図を多用する

- ▶ 「ものの考え方」としての統計的思考方法の重要性を伝えたい

⇒「データに基づく意思決定」の重要性を訴求

課題の把握から問題解決の方法を探るPPDACサイクルの重要性を訴求

セミナーの実施について

(株)Rejouiの企画・運営により
2/20(土), 21(日)の両日10:00~17:30(70分×5コマ)
総務省統計研究研修所(最寄り駅:JR西国分寺駅)で実施
オンラインでも同時開催

両日ともに

集合研修:30名(オンラインに振替になる可能性も有り)

オンライン研修:100名

※参加希望者多数の場合は、抽選とします。後日動画配信を予定しています

詳細はイベント公式サイトに掲載

<https://datascience-seminar.jp/>

Thank you for your attention.



総務省統計研究研修所

