

# 国立10大学における データサイエンス教育の内容分類



TOKYO METROPOLITAN UNIVERSITY

東京都立大学

椿本 弥生 (Tsubakimoto, Mio)

東京都立大学 大学教育センター  
mio@tmu.ac.jp



## 背景

- 大学におけるデータサイエンス（以下、DS）教育の必要性
  - 低年次生（学部1, 2年生）から受講できるDS授業
- 
- 理系のみならず文系も対象
  - DSは、主に理工系分野の学問内容に基づく
    - 文系学生：授業内容や方法に何らかの工夫が必要
    - 理系学生：何をどこまで教えられるかの判断は難しい



# 目的

- 大学学部の学部1, 2年生を主な対象とした  
基礎的なDS授業のシラバスの内容分類
- 文理それぞれのDS授業における  
内容の共通性や独自性を探索的に発見
- 大学がWebで外部公開しているシラバスの  
テキストデータを収集し、テキストマイニング

## 方法 データ抽出の手続き

- 国立大学86校<sup>[1]</sup> → 指定国立大学法人10校<sup>[2]</sup>
- 外部向けWebシラバスから手動で検索・抽出
  - 2022年度内に開講
  - 授業名、概要、目的・目標、内容・計画、方法のいずれかに“データサイエンス”の単語を含む
  - 主に学部1、2年生向け、一般教養科目や全学共通教育科目

[1] 文部科学省（2021），令和3年度全国大学一覧

[2] 文部科学省（2021），第4期中期目標機関における指定国立大学法人の指定について



## 方法 内容分析

- 内容分析 : KH Coder (Version 3.Alpha.11a)
- 形態素解析 : MeCab (KH Coder内蔵)
- 手順 :
  1. 形態素解析
  2. 品詞ごとの単語の分割状況、出現頻度の確認
  3. データクレンジング
  4. 以上を繰り返す

## • 強制的に使用した語

データサイエンス、R言語、R、Python、機械学習、情報科学、問題解決、ニューラルネットワーク、AI、人工知能、Anaconda、GoogleColaboratory、MapReduce、OS、PC、Scikit-Learn、Pandas、NumPy、SQL、グループワーク、教師あり学習、教師なし学習、t検定、t分布、z検定、利活用、Excel、可視化、視覚化、図表

## • 強制的に使用しなかった語

第1回、第2回、第3回、第4回、第5回、第6回、第7回、第8回、第9回、第10回、第11回、第12回、第13回、第14回、第15回、(1)、(2)、(3)、(4)、(5)、I、II、Part、.、(、)、,、/、))、):、~、~、):、-、:、講義、題目

※ 複合語の設定プロトコルについて補足 (レビュアーからのコメント)

# 結果 授業数と授業形式

- 10大学の総授業数：31 ( $Md = 3$ ,  $SD = 1.45$ ,  $Min = 1$ ,  $Max = 6$ )
  - 総合大学 7、理系大学 2、文系大学 1

- いずれの大学も、講義と演習を組み合わせた「講義・演習形式」が多数
- 理系大学は、演習中心の割合が高い

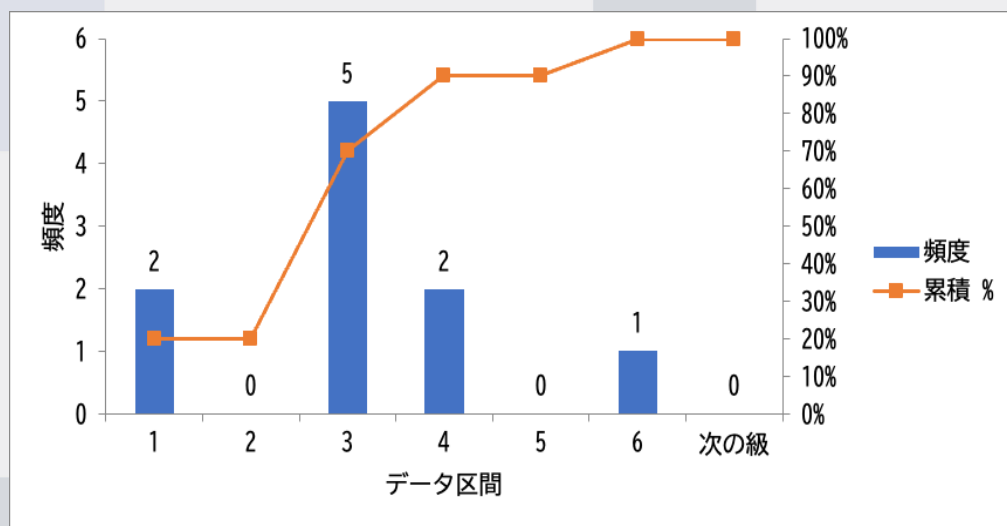


表 1 授業内容の分類

	講義	演習	講義・演習	計
総合大学	6	1	12	19
理系大学	3	2	5	10
文系大学	0	0	2	2
授業数計	9	3	19	31

図 1 授業数の分布

# 結果 内容・計画×分野の共起ネットワーク

• 「内容」「計画」に含まれる語 × 分野（総合、理系、文系）との関係

## • 文系

- 実際に「プログラミング」して計算する
- AIがどのように「社会」で「活用」されているか学ぶ

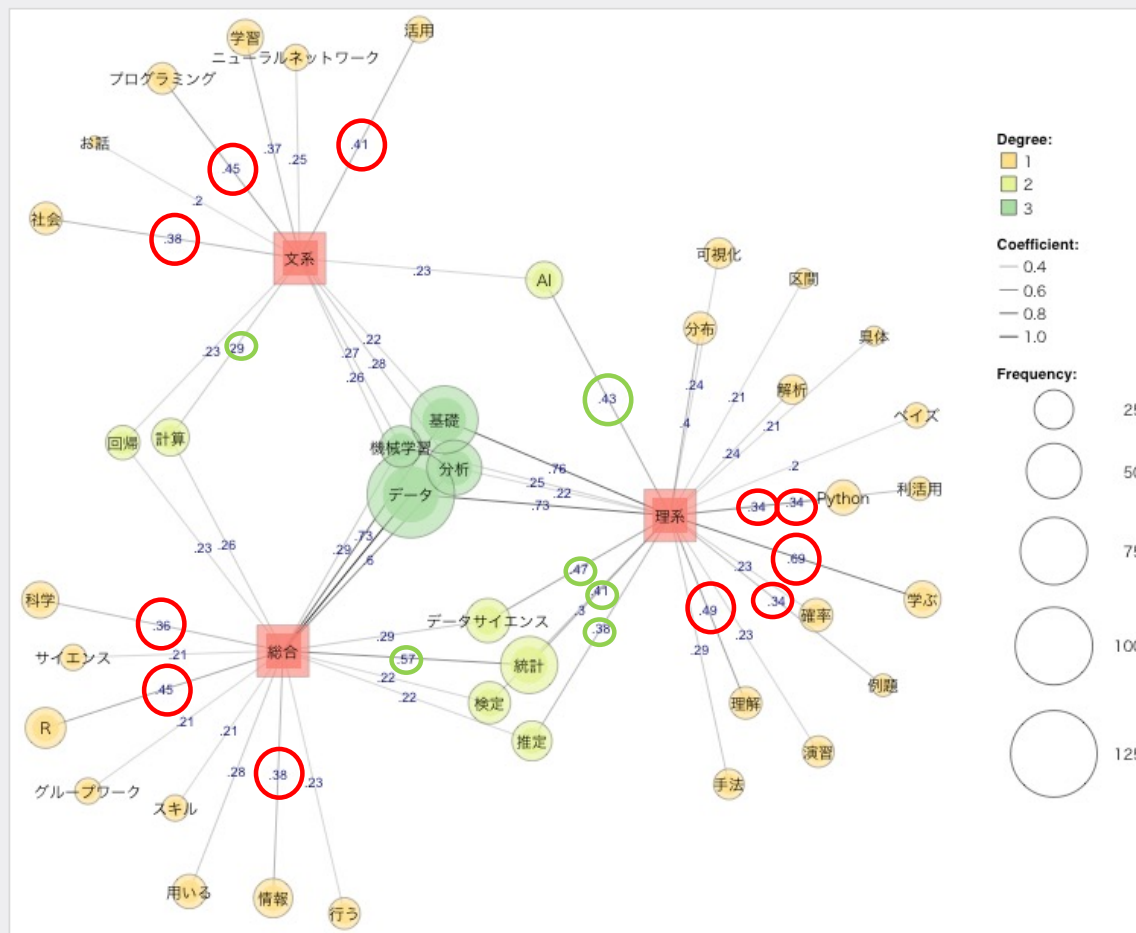
## • 理系

- 「Python」を用いた「例題」を通し、統計学の「利活用」を「学ぶ」「理解」する

## • 総合

- 「R」を用いた統計「科学」に加え「情報」表現や「情報」関連法令も含めた社会的・学術的背景も学ぶ

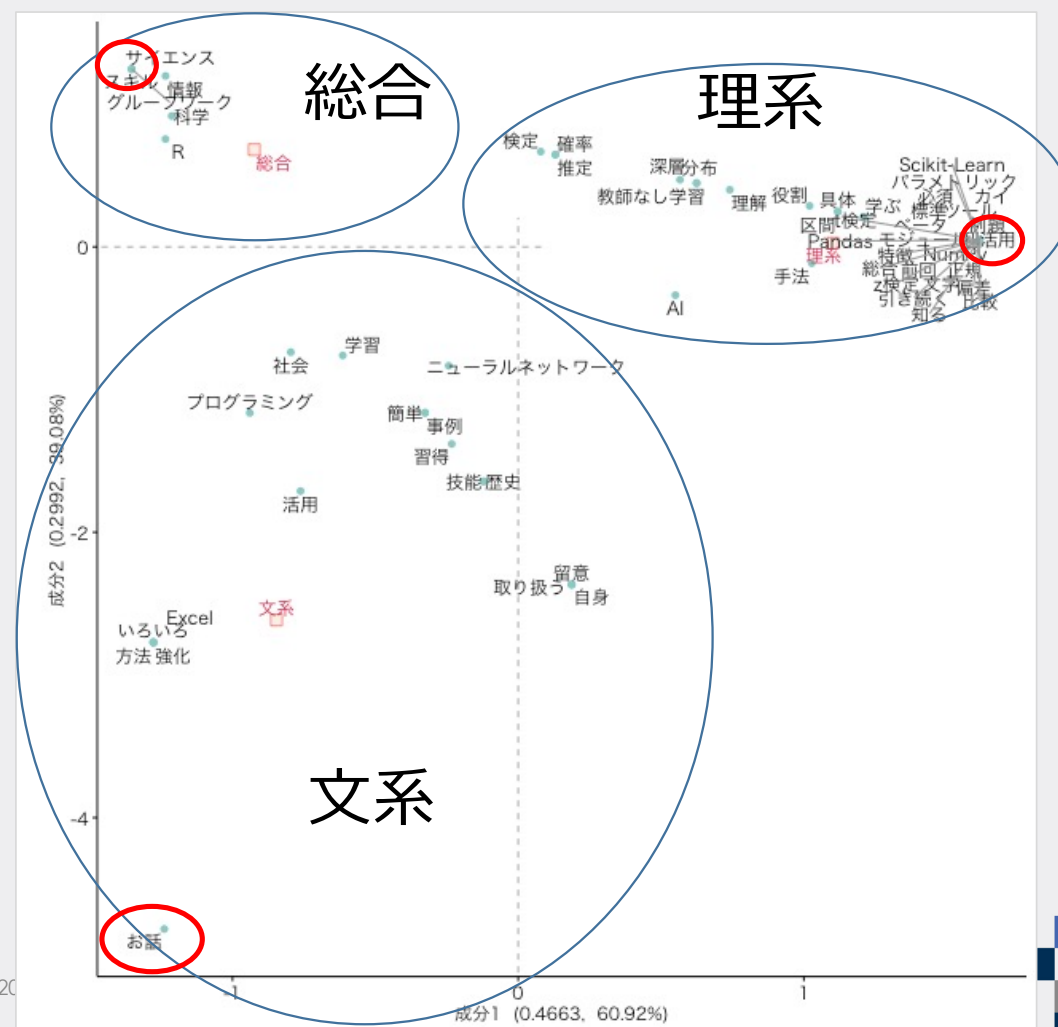
2022/11/17





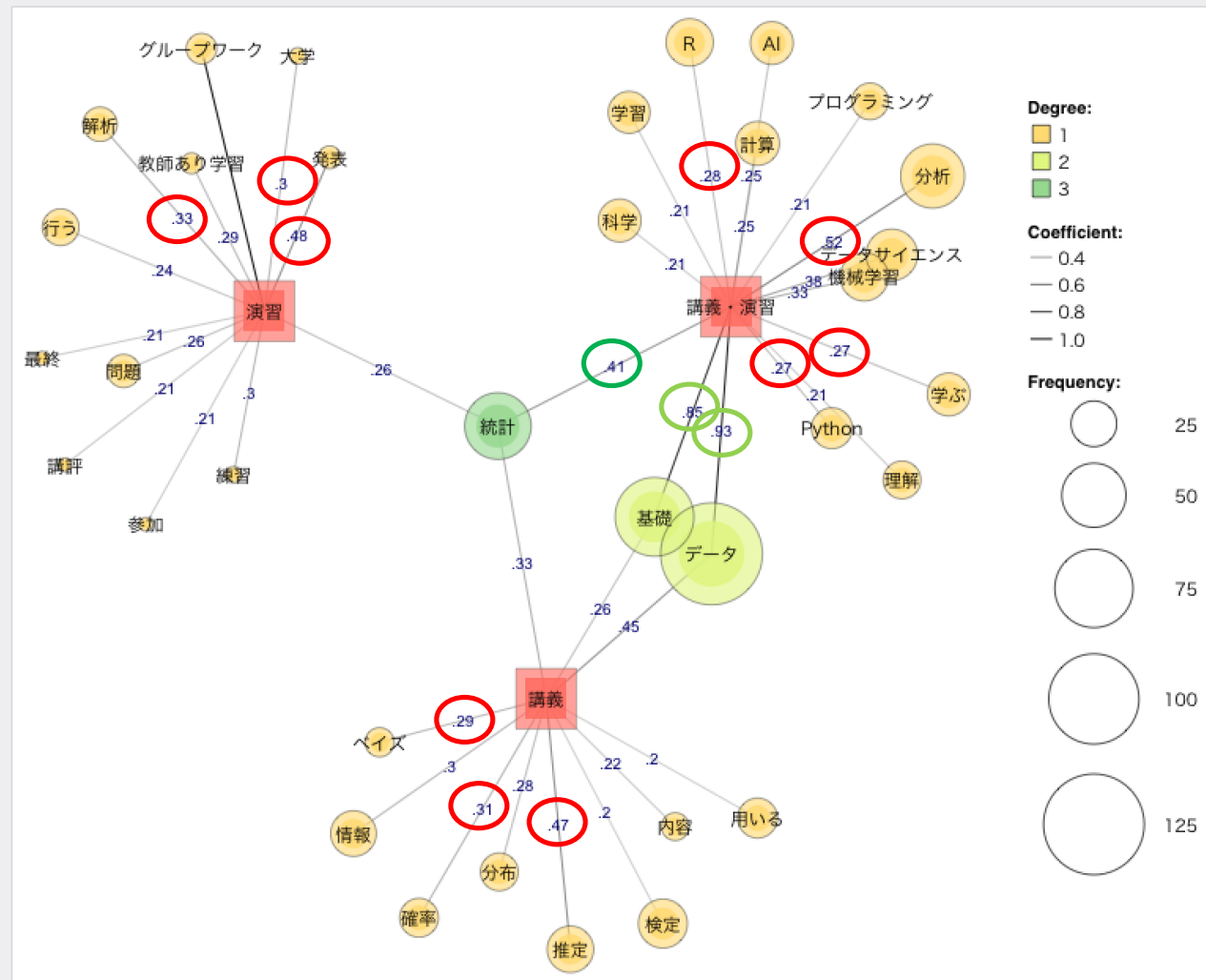
# 結果 内容・計画×分野の対応分析

- 分野に最も特徴的な単語は？
- 文系 **心理的負担の軽減**
  - 「お話」
    - cf. 簡単な統計のお話、ディープラーニングのお話
- 理系 **統計とプログラミング**
  - 「偏差」「回帰」統計用語
  - 「Pandas」ライブラリ
  - 「例題」「知る」授業内容
- 総合 **演習形式で学ぶ**
  - 「サイエンス」の「スキル」を「グループワーク」で学ぶ



# 結果 内容・計画×授業方法の共起ネットワーク

- 3タイプに共通「統計」
- 【講義・演習】と【講義】共通「基礎」「データ」
- 【講義】
  - 確率・推定、ベイズの定理
- 【講義・演習】
  - RやPythonを用いて分析を学ぶ
- 【演習】
  - グループで解析 → 発表



## まとめ・本研究の限界・今後の課題

- 指定国立大学10校のDS授業シラバスの内容分析
- 分野や授業方法による、授業内容の共通点・差異
  
- 手動で収集、学外向け・限定的内容のシラバス → データの漏れ？
- 今回の結果は限定的
  
- 公立大、私大、科目区分（全学共通科目、学部など）も含め、より大規模にデータを収集し同様の分析を試みる
- ただし、各大学のシラバスWebサイトの構成は多様であるため、効率的な収集手法を検討する必要