

情報Ⅰから情報Ⅱへ データサイエンス実践

日出学園中学校・高等学校
武善 紀之

<https://www.hinode.ed.jp/share/takeyoshi/>



自己紹介

自己紹介

名前 武善紀之(たけよし のりゆき)

所属 私立 日出学園中学校・高等学校 教諭
所在地:千葉県市川市 ※幼稚園・小学校併設

教科 情報科 11年目(+数学,公民,技術)

現在 高2担任・パソコン部顧問

年齢 33歳

出身 筑波大学情報学群情報メディア創成学類

好きなもの ペンギン



<https://www.hinode.ed.jp/share/takeyoshi/>

「情報科」関連の主な兼職

• 教科書執筆

- 情報科「新編情報Ⅰ」「情報Ⅰ Step Forward!」「情報Ⅱ」(東京書籍)

• 予備校

- 「ベーシックレベル情報Ⅰ」講師(スタディサプリ)

• 教育番組

- NHK高校講座「情報Ⅰ」監修講師(NHK)

• 教員研修

- GIGA スクールにおける学びの充実「高等学校情報教員指導力向上事業」(文部科学省)
- GIGA スクール「情報Ⅱ」高等学校情報科等強化によるデジタル人材の供給体制整備支援事業(文部科学省)

スタディサプリ



NHK 高校講座



武善 紀之
たけよし のりゆき
日出学園中学校・高等学校教諭



IT活用セミナーは3年前にも登壇

【オンライン高校IT活用セミナー】

ICT機器の整備と活用—小中学校1人1台環境整備を受けて

2021年01月16日(土) 13時~15時50分

情報Ⅰ,Ⅱにつなげる ”情報科”統計教育

日出学園中学校・高等学校
武善 紀之





⇨「Takeyoshi Noriyuki」で検索すると出ます

http://high.hinode.ed.jp/share/takeyoshi/n_takeyoshi.html

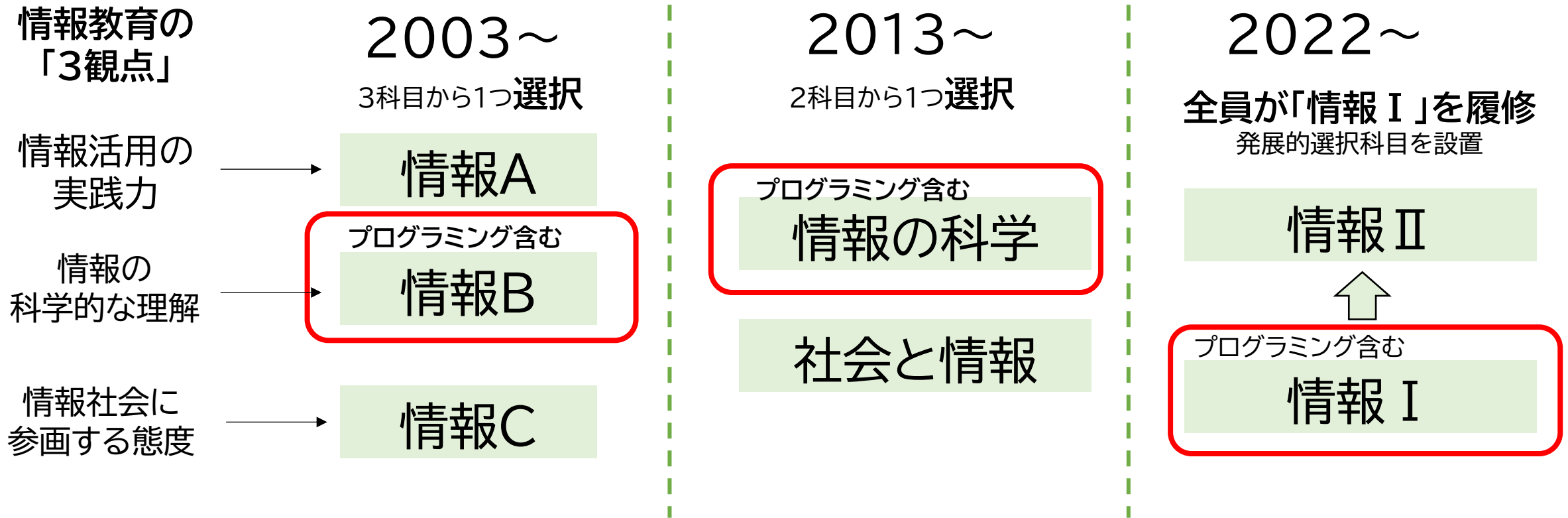
(2021) 情報Ⅰ,Ⅱにつなげる“情報科”統計教育

(2024) 情報Ⅰから情報Ⅱへ データサイエンス実践

Agenda.

-  情報Ⅰ／情報Ⅱのデータサイエンス
(学習指導要領より)
-  情報Ⅱにおける実践
-  情報Ⅰにおける実践
-  データサイエンスを扱う価値3つ(個人的に)

情報Ⅰと情報Ⅱに関するざっとしたおさらい



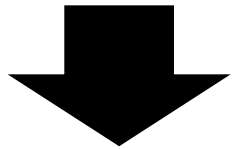
×新教科「情報Ⅰ」では扱う内容が大幅に増加！

○「プログラミング」や「シミュレーション」については、
今まで「選択」であった内容が「必修」になった。

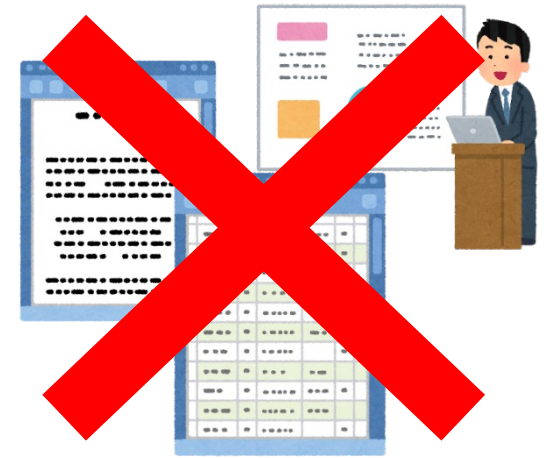
新学習指導要領「情報Ⅰ」

「情報Ⅰ」のコンセプト

1. コンピュータを活用して、**問題解決**ができるようになること
2. **情報科学(Computer Science)**をきちんと学ぶこと



- ◆第1章 情報社会の問題解決
- ♥第2章 コミュニケーションと情報デザイン
- ♣第3章 コンピュータとプログラミング
- ♠第4章 情報通信ネットワークとデータの活用



コンピュータを活用して問題解決ができるようになる

「情報デザイン」「データサイエンス」「プログラミング」は、
問題解決の三種の神器

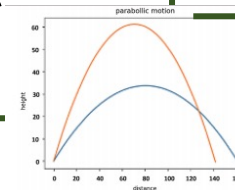


魅力的な
学園祭を
作り上げよう！

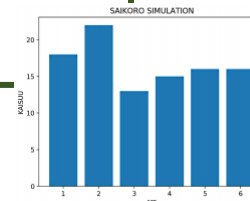
魅力的なポスターを作って集客しよう！
情報デザイン



幾らで売れば、一番利益が出るだろう？
プログラミング



過去の来客数を分析して、傾向を判断！
データサイエンス



最高の学園祭
を実現！

これを受けて情報Ⅱでは…？

おおざっぱな「情報Ⅱ」のイメージ

データサイエンスの学習内容は、

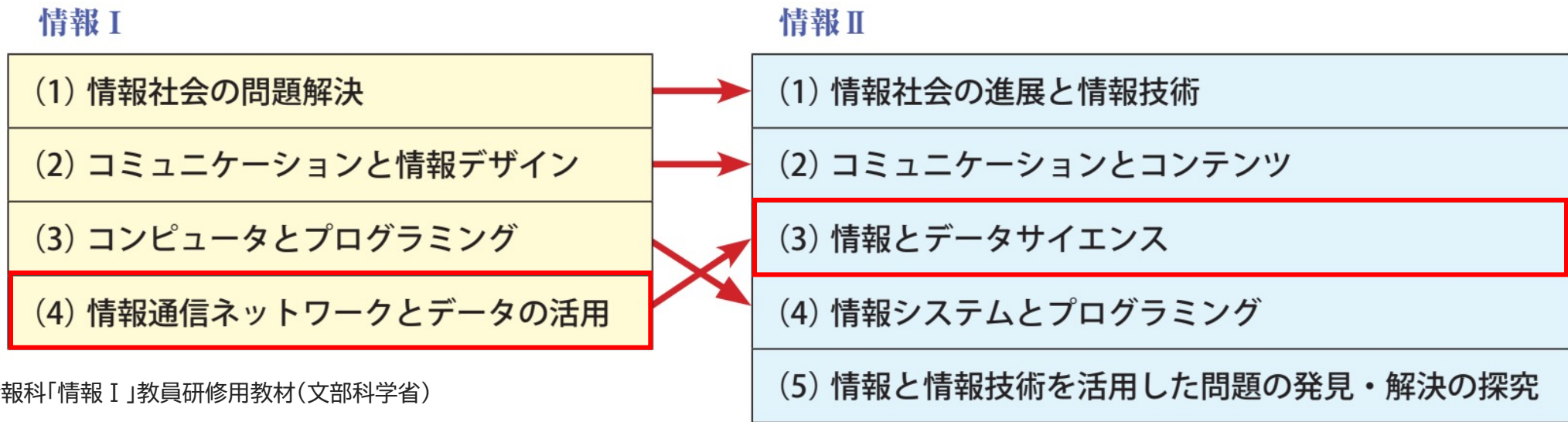
大雑把にまとめると**3つ**

1. 自分達でアンケート集計や分析を行う内容 →主に、「情報Ⅰ」
2. データベースや情報システムに関する内容 →両方
3. **大規模データの分析,可視化に関する内容** →主に「情報Ⅱ」

ともかく大規模！
10次元以上とか当たり前。

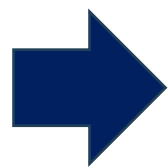
	01_睡眠	02_身の回りの用事	03_食事	04_通勤・通学	05_仕事	06_学業	07_家事	10_買い物	11_移動(通勤・通学を除く)	12_テレビ・ラジオ・新聞・雑誌	13_休養・くつろぎ	14_学習・自己啓発・訓練(学業以外)	15_趣味・娯楽	16_スポーツ
01_北海道	467	72	74	62	41	355	5	3	19	25	166	42	63	18
02_青森県	469	71	90	53	34	368	0	0	13	79	128	19	74	30
03_岩手県	458	69	76	53	47	453	5	3	8	31	105	42	59	14
04_宮城県	448	75	80	72	44	373	3	3	7	16	93	29	110	53
05_秋田県	467	61	84	45	29	373	4	6	13	19	136	32	114	34
06_山形県	477	74	83	52	43	377	4	5	6	40	135	38	58	29
07_福島県	436	80	87	64	41	438	7	2	15	23	146	36	40	25
08_茨城県	456	88	76	71	31	387	6	6	18	24	133	41	48	35
09_栃木県	472	66	82	69	46	394	13	12	9	24	131	13	72	20
10_群馬県	457	74	78	69	41	398	7	7	12	31	125	36	67	22
11_埼玉県	439	83	91	80	31	401	5	3	12	28	134	55	41	13
12_千葉県	437	71	85	85	70	367	8	3	12	15	170	54	33	17
13_東京都	451	84	86	80	33	387	5	7	11	16	124	55	56	36
14_神奈川	452	89	86	77	32	399	2	8	14	39	108	23	78	9
15_新潟県	475	71	74	50	13	381	7	11	8	49	143	49	62	21
16_富山県	475	66	75	66	68	376	3	4	10	35	125	28	64	15
17_石川県	460	69	78	51	48	395	3	1	16	16	135	31	79	43
18_福井県	445	87	75	56	36	450	3	4	10	12	126	54	45	25
19_山梨県	453	75	78	62	28	429	5	1	16	30	94	60	66	14
20_長野県	463	62	91	72	43	402	10	1	6	26	150	30	53	17
21_岐阜県	458	73	78	69	38	418	13	6	9	45	101	36	57	23
22_静岡県	452	82	81	62	49	390	8	3	11	21	136	41	54	23
23_愛知県	459	93	79	71	72	374	4	1	12	25	123	44	51	13
24_三重県	437	66	75	81	87	358	3	2	6	21	171	42	41	23
25_滋賀県	462	81	83	60	47	398	5	4	6	12	144	46	47	27
26_京都府	478	65	83	53	23	360	17	10	14	27	159	34	53	15
27_大阪府	430	88	84	70	65	366	8	6	16	23	147	56	51	22
28_兵庫県	465	79	87	69	45	351	3	6	12	33	161	40	48	19
29_奈良県	446	72	86	107	10	416	7	15	9	41	75	36	100	13
30_和歌山	467	58	81	71	31	403	7	6	14	37	139	31	45	23
31_鳥取県	464	69	72	55	55	408	9	4	13	37	138	38	27	31
32_島根県	449	68	75	66	47	440	1	1	8	14	152	51	37	19
33_岡山県	424	88	82	77	40	405	4	2	21	14	135	49	40	11
34_広島県	450	64	80	67	27	433	18	6	7	20	132	28	71	17
35_山口県	432	70	81	67	63	384	7	5	20	25	133	34	63	32
36_徳島県	451	76	77	52	65	401	3	3	8	31	111	61	64	18
37_香川県	450	61	70	67	60	380	5	6	11	20	110	60	60	10

指導要領の言葉を比較してみる



図版:高等学校情報科「情報Ⅰ」教員研修用教材(文部科学省)

情報Ⅰ(必修) ※学習指導要領解説より
質的データと量的データ,データの整理(データクリーニングなど),欠損値や外れ値の扱い方,データの様々な可視化,テキストマイニングオープンデータの活用,散布図行列,相関と交絡因子,単回帰分析,統計的仮説検定



情報Ⅱ(選択科目) ※学習指導要領解説より
選択バイアス,情報バイアス,機械学習,訓練データ,過剰適合,重回帰分析,モデルの調整,主成分分析,クラスタリング,文字認識,画像認識,人工知能

とてもじゃないけど、できない？

機械学習をまとめると

機械学習

機械が、データからパターンや規則を学習し、予測や判断を行う技術。
AI(人工知能)を実現する技術の1つ。

正解データがある

正解データがない

教師あり学習

数値の予測

回帰

Regression

- ・線形回帰

データの分類

クラス分類

Classification

- ・ロジスティック回帰
- ・決定木

教師なし学習

主要な特徴を
残す・取り出す

次元削減

Dimensionality
Reduction

- ・主成分分析

データの
グループ化

クラスタリング

Clustering

- ・k平均法

やっぱり、とてもじゃないけど、できない？

実は「情報Ⅰ」より「情報Ⅱ」の方が
面白い／簡単？

ここまで無理だと諦めがつく

ともかく大規模！
10次元以上とか当たり前。

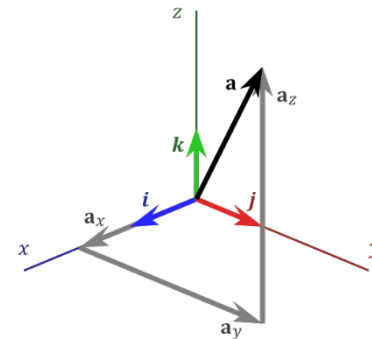
	01_睡眠	02_身の回りの用事	03_食事	04_通勤・通学	05_仕事	06_学業	07_家事	10_買い物	11_移動(通勤・通学を除く)	12_テレビ・ラジオ・新聞・雑誌	13_休養・くつろぎ	14_学習・自己啓発・訓練(学業以外)	15_趣味・娯楽	16_スポーツ
01_北海道	467	72	74	62	41	355	5	3	19	25	166	42	63	18
02_青森県	469	71	90	53	34	368	0	0	13	79	128	19	74	30
03_岩手県	458	69	76	53	47	453	5	3	8	31	105	42	59	14
04_宮城県	448	75	80	72	44	373	3	3	7	16	93	29	110	53
05_秋田県	467	61	84	45	29	373	4	6	13	19	136	32	114	34
06_山形県	477	74	83	52	43	377	4	5	6	40	135	38	58	29
07_福島県	436	80	87	64	41	438	7	2	15	23	146	36	40	25
08_茨城県	456	88	76	71	31	387	6	6	18	24	133	41	48	35
09_栃木県	472	66	82	69	46	394	13	12	9	24	131	13	72	20
10_群馬県	457	74	78	69	41	398	7	7	12	31	125	36	67	22
11_埼玉県	439	83	91	80	31	401	5	3	12	28	134	55	41	13
12_千葉県	437	71	85	85	70	367	8	3	12	15	170	54	33	17
13_東京都	451	84	86	80	33	387	5	7	11	16	124	55	56	36
14_神奈川県	452	89	86	77	32	399	2	8	14	39	108	23	78	9
15_新潟県	475	71	74	50	13	381	7	11	8	49	143	49	62	21
16_富山県	475	66	75	66	68	376	3	4	10	35	125	28	64	15
17_石川県	460	69	78	51	48	395	3	1	16	16	135	31	79	43
18_福井県	445	87	75	56	36	450	3	4	10	12	126	54	45	25
19_山梨県	453	75	78	62	28	429	5	1	16	30	94	60	66	14
20_長野県	463	62	91	72	43	402	10	1	6	26	150	30	53	17
21_岐阜県	458	73	78	69	38	418	13	6	9	45	101	36	57	23
22_静岡県	452	82	81	62	49	390	8	3	11	21	136	41	54	23
23_愛知県	459	93	79	71	72	374	4	1	12	25	123	44	51	13
24_三重県	437	66	75	81	87	358	3	2	6	21	171	42	41	23
25_滋賀県	462	81	83	60	47	398	5	4	6	12	144	46	47	27
26_京都府	478	85	83	53	23	360	17	10	14	27	159	34	53	15
27_大阪府	430	88	84	70	65	366	8	6	16	23	147	56	51	22
28_兵庫県	465	79	87	69	45	351	3	6	12	33	161	40	48	19
29_奈良県	446	72	86	107	10	416	7	15	9	41	75	36	100	13
30_和歌山	467	58	81	71	31	403	7	6	14	37	139	31	45	23
31_鳥取県	464	69	72	55	55	408	9	4	13	37	138	38	27	31
32_島根県	449	68	75	66	47	440	1	1	8	14	152	51	37	19
33_岡山県	424	88	82	77	40	405	4	2	21	14	135	49	40	11
34_広島県	450	64	80	67	27	433	18	6	7	20	132	28	71	17
35_山口県	432	70	81	67	63	384	7	5	20	25	133	34	63	32
36_徳島県	451	76	77	52	65	401	3	3	8	31	111	61	64	18
37_香川県	450	61	70	67	20	300	5	0	11	20	110	20	60	10

人工知能(AI)やコンピュータを
活用しないと、無理！

(情報 I はギリギリ「できてしまう」範囲だった)
コンピュータに頼らざるを得ない。

※僕のイメージ

高校数学のベクトル
(3次元まで)



大学の線形代数
(n次元空間)

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

※図はWikipediaより

指導要領もよく読むと……

(情報Ⅱ学習指導要領解説)

ここでは、数学科における学習内容と関連する部分も含むが、**数学や統計学の専門的な内容に深入りすることなく、可視化やソフトウェアによる処理の結果を基に、その概念を理解するようにする。**

(実際に指導してみると……)

- 情報Ⅰより、**ソフトウェアの操作が簡単**(コマンド入力で終わり)
- 情報Ⅰより、**数学感が薄い**(うわ、苦手だった奴……がない)
- 情報Ⅰより、**扱えるデータが多様**(みんな全く違うテーマで楽しめる)

(生徒も)

- 情報Ⅱは実践的なものが多かったので**楽しい**。
- たくさんのデータをまとめることは数学などでは難しいので、それらをいっきにまとめられるのは**とても便利だと思った**。

僕がイメージしているゴール設計

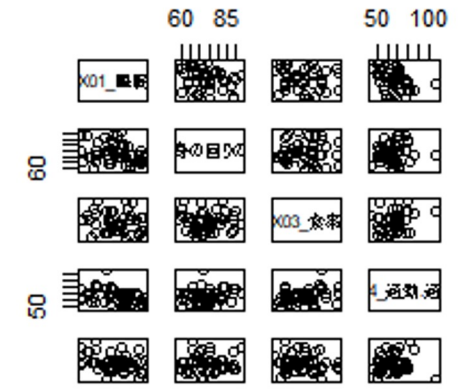
多次元のデータ

01.睡眠	02.身の回り	03.食事	04.通勤・通学	05.仕事	06.学業	07.家事	08.買い物	09.移動・通	10.趣味・娯楽	11.テレビ	12.休養・く	13.学習・自	14.スポーツ	15.その他
01.睡眠	02.身の回り	03.食事	04.通勤・通学	05.仕事	06.学業	07.家事	08.買い物	09.移動・通	10.趣味・娯楽	11.テレビ	12.休養・く	13.学習・自	14.スポーツ	15.その他

情報 I

2つずつで
頑張ってみる

	01.睡眠	02.身の回り	03.食事	04.通勤・通学	05.仕事	06.学業	07.家事	08.買い物	09.移動・通	10.趣味・娯楽	11.テレビ	12.休養・く	13.学習・自	14.スポーツ	15.その他
01.睡眠															
02.身の回り	-0.2308														
03.食事	-0.129	0.04927													
04.通勤・通学	-0.3563	0.19467	0.28387												
05.仕事	-0.3274	0.08675	-0.1459	0.02845											
06.学業	-0.3835	-0.1301	-0.2055	-0.0192	-0.0992										
07.家事	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732									
10.買い物	0.26421	0.02212	0.0845	0.12755	-0.4664	-0.239									
11.移動・通	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163									
12.テレビ	0.35748	-0.1084	0.11338	-0.0222	-0.2739	-0.0922									
13.休養・く	0.09791	-0.1746	0.04584	-0.1438	0.2955	-0.4069									
14.学習・自	-0.1175	0.37528	-0.0702	0.03141	0.02206	-0.2363									
15.趣味・娯	0.12724	-0.1379	0.11829	-0.052	-0.3413	-0.2828									
16.スポー	0.03883	-0.0809	-0.0632	-0.2982	0.01111	-0.1535									
20.その他	0.10732	-0.0867	-0.2132	-0.3353	-0.1576	0.01509									

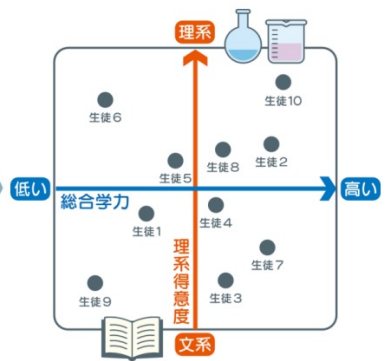
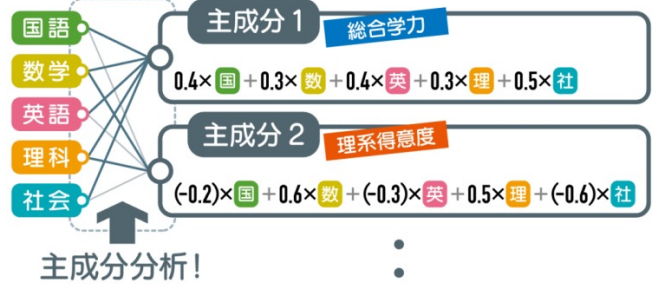


相関行列

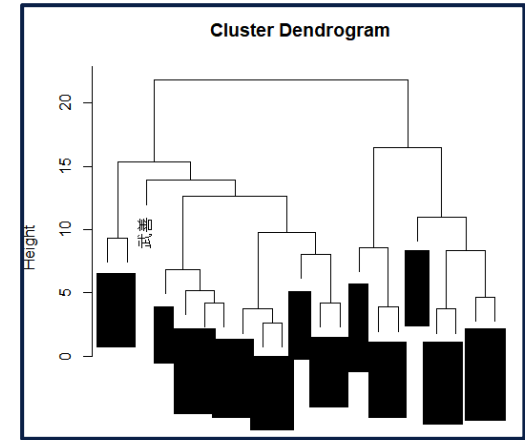
散布図行列

情報 II そのまま見る!

主成分分析で、次元を圧縮しよう



クラスタリングで、グループ化しよう



etc...

情報Ⅰ／情報Ⅱ解説動画 プロジェクト

文科省「情報Ⅰ」「情報Ⅱ」解説動画 <https://www.nttls-edu.jp/joho/>

- ・ データサイエンス領域で情報Ⅰ、情報Ⅱともに関わった
- ・ 「これぞ情報Ⅰ」「これぞ情報Ⅱ」と言える内容にはなったと思われる

情報Ⅰ



情報Ⅱ



以降で、

情報Ⅰと情報Ⅱにおけるデータサイエンスの実践を紹介

目標

情報Ⅱ、案外出来そうだな

情報Ⅱ、やってみたくなる！

Agenda.

- 1 情報Ⅰ / 情報Ⅱのデータサイエンス(学習指導要領より)
- 2 情報Ⅱにおける実践**
- 3 情報Ⅰにおける実践
- 4 データサイエンスを扱う価値3つ(個人的に)

情報Ⅱの開設状況(2024年度が初年度)

4年(高1)

必履修2単位

情報Ⅰ

5年(高2)

選択2単位【文理共通】

人間と機械[※]

※学校設定科目(公民)

6年(高3)

選択2単位【理系】

情報Ⅱ

選択2単位【文理共通】

実践情報Ⅰ

※「人間と機械」については先週、全高情研で発表。

「倫理」をベースとして、「人間と機械の関係性」を考える科目。

発表資料は既に個人Webへ掲載済。



但し、実際は9年目とも言える。

- 元々、高3の選択科目として「情報の科学」を開講(初年度のみ社会と情報)。
- 科学の頃は重複も多いので、1学期科学分野、2学期を卒業研究・制作としていた。
- 情報Ⅱ関連の執筆でも、このコマを実証実験として活用できていたのは大きい。

年度	担当科目
2016年	社会と情報(4),社会と情報(2),中1数学(5),高3数学(6)
2017年	社会と情報(9),情報の科学(4),デジタルコンテンツ演習(4)
2018年	社会と情報(10),情報の科学(2),デジタルコンテンツ演習(4)
2019年	社会と情報(9),情報の科学(2)
2020年	社会と情報(10),情報の科学(2),倫理(2),高2探究(1)
2021年	社会と情報(9),情報の科学(2),倫理(2),中2総合(4),高2探究(1)
2022年	情報Ⅰ(5),社会と情報(6),情報の科学(2),倫理(2)
2023年	情報Ⅰ(10),情報の科学(2),人間と機械(2),社会と情報(1)
2024年	情報Ⅰ(10),情報Ⅱ(2),実践情報Ⅰ(2),人間と機械(2)

情報Ⅱの開設状況

4年(高1)

必修2単位

情報Ⅰ

5年(高2)

選択2単位【文理共通】

人間と機械[※]

※学校設定科目(公民)

6年(高3)

選択2単位【理系】

情報Ⅱ

選択2単位【文理共通】

実践情報Ⅰ

- 履修者45名(!) → 情報科初のTT(卒業生支援)
- 受験科目が少ない理系生徒の必修選択枠

情報Ⅱ 2024年度年間指導計画(?)

(1学期)

- Webページの作成(HTML & CSS & JavaScript) [5h]
- 人工知能入門(OpenCVによる画像認識)[2h]
- TWICE PLAN「次世代技術探究ワーク」[5h]
- クラスタリングによる生徒分類[1h]
- メタセコイアを用いた3DCG入門[1h]
- 主成分分析[3h]
- 授業内小テスト&返却[2h]

(2学期)

- Adobe CCを用いたコンテンツ制作[3h]
- 3Dプリンタを活用した3D造形[3h]
- 卒業研究・卒業制作[10h]

データサイエンスだけ抜き出すと……

2h

人工知能入門 モデルを用いた画像認識
人工知能(AI)とデータの関係性って何？

情報II

モデルを用いた
画像認識

武善 紀之



1h

クラスタリングで似ている友達を探せ
仕組みの概念的な理解を試してみる

情報II

クラスタリング

武善 紀之



3h

主成分分析で「好きなもののマップ」を作ろう
実際にデータサイエンスを使ってみる！

情報II

主成分分析による
次元縮約

武善 紀之



データサイエンスだけ抜き出すと……

2h

人工知能入門 モデルを用いた画像認識
人工知能(AI)とデータの関係性って何？

情報II

モデルを用いた
画像認識

武善 紀之



1h

クラスタリングで似ている友達を探せ
仕組みの概念的な理解を試してみる

情報II

クラスタリング

武善 紀之



3h

主成分分析で「好きなもののマップ」を作ろう
実際にデータサイエンスを使ってみる！

情報II

主成分分析による
次元縮約

武善 紀之



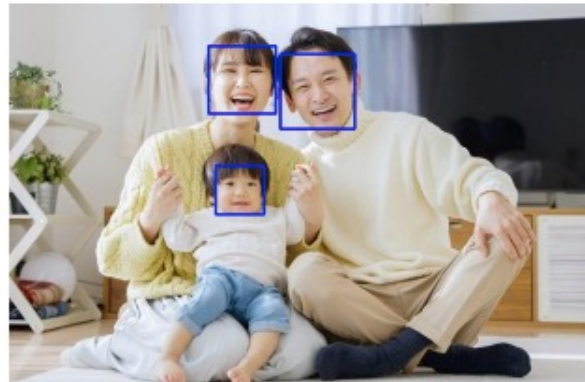
①人工知能入門 モデルを用いた画像認識

<https://www.youtube.com/watch?v=EQAmucowTsc>

この動画でやってみること

学習については別動画で!

学習済みのモデルを使って、
顔の検出・ぼかしに挑戦してみよう!

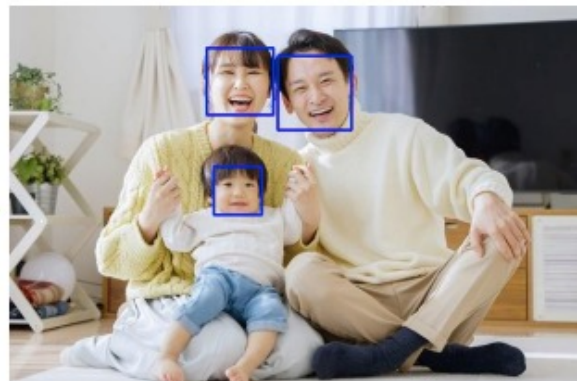
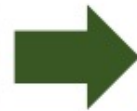


①人工知能入門 モデルを用いた画像認識

Google Colaboratoryを使用

<https://www.youtube.com/watch?v=EQAmucowTsc>

```
01 import cv2
02 from google.colab.patches import cv2_imshow
03 img = cv2.imread('syugou.jpg')
04 cv2_imshow(img)
05
06 face_cascade = cv2.CascadeClassifier('haarcascade_frontalface_alt.xml')
07 faces = face_cascade.detectMultiScale(img)
08
09 for(x,y,w,h) in faces:
10     face = img[y:y+h,x:x+w]
11     face = cv2.blur(face,(55,55))
12     img[y:y+h,x:x+w]=face
13 if len(faces)!=0:
14     cv2_imshow(img)
```



①人工知能入門 モデルを用いた画像認識

<https://www.youtube.com/watch?v=EQAmucowTsc>

なぜ、これが
AI・データサイエンスなのか？

①人工知能入門 モデルを用いた画像認識

<https://www.youtube.com/watch?v=EQAmucowTsc>

1番最初に実施 → 「データサイエンス」と「AI」の接続

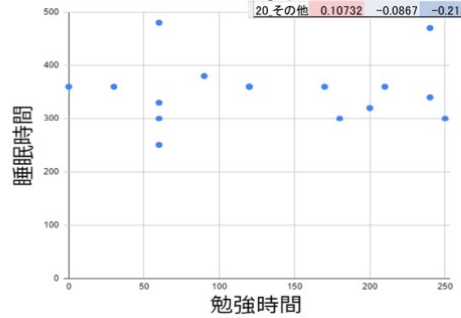
「顔認識AI」



実は「データサイエンス」単元と深い関わりを持つ

データサイエンスというと…?

01_睡眠	06_学業
467	355
469	368
458	453
448	373
467	373
477	377
436	438
456	387
479	394



01_睡眠	02_身の回り	03_食事	04_通勤・通学	05_仕事	06_学業	07_家事	10_買い物(通勤・通学・ラ)
01_睡眠	-0.2308						
02_身の回り	-0.129	0.04927					
03_食事	-0.3563	0.19467	0.28387				
04_通勤・通学	-0.3274	0.08675	-0.1459	0.02845			
05_仕事	-0.3835	-0.1301	-0.2055	-0.0192	-0.0992		
06_学業	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732	
07_家事	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163	-0.0613
10_買い物(通勤・通学・ラ)	0.26421	0.02212	0.0845	0.12755	-0.4664	-0.239	0.40576
移動	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163	-0.0613
テレビ	0.35748	-0.1084	0.11338	-0.0222	-0.2739	-0.0922	0.00931
13_休養	0.09791	-0.1746	0.04584	-0.1438	0.2955	-0.4069	-0.0013
14_学習	-0.1175	0.37528	-0.0702	0.03141	0.02206	-0.2363	-0.2223
15_趣味	0.12724	-0.1379	0.11829	-0.052	-0.3413	-0.2828	-0.0363
16_スポーツ	0.0383	-0.0809	-0.0632	-0.2982	0.01111	-0.1535	-0.0966
20_その他	0.10732	-0.0867	-0.2132	-0.3353	-0.1576	0.01509	0.19406



①人工知能入門 モデルを用いた画像認識

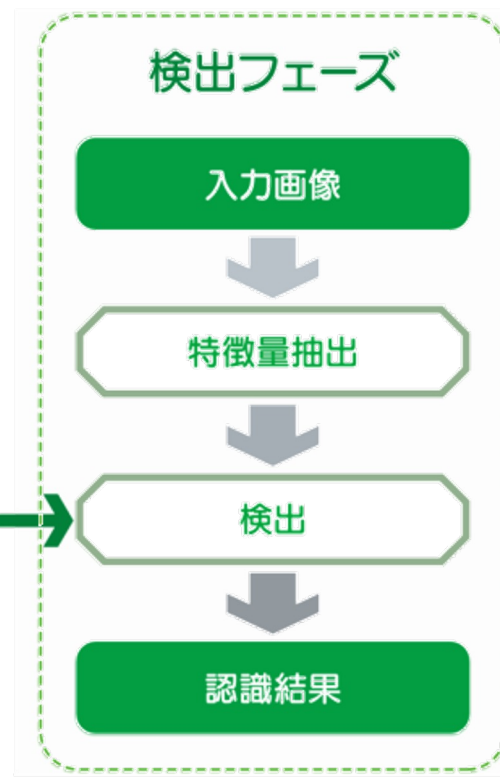
<https://www.youtube.com/watch?v=EQAmucowTsc>

1番最初に実施 → 「データサイエンス」と「AI」の接続

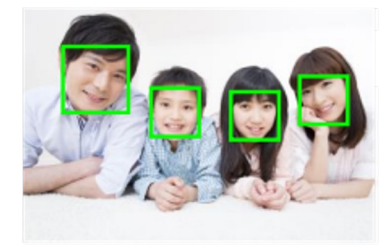
「データ」から学習し、作ったモデルで検出



顔って、
こんなものだ!



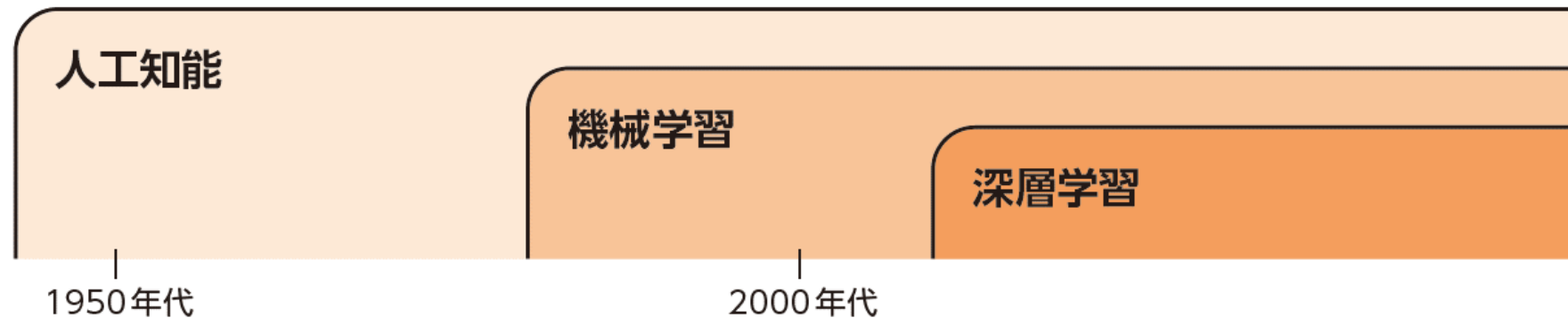
顔っぽいのは、
ここだ!



第3章 情報とデータサイエンス

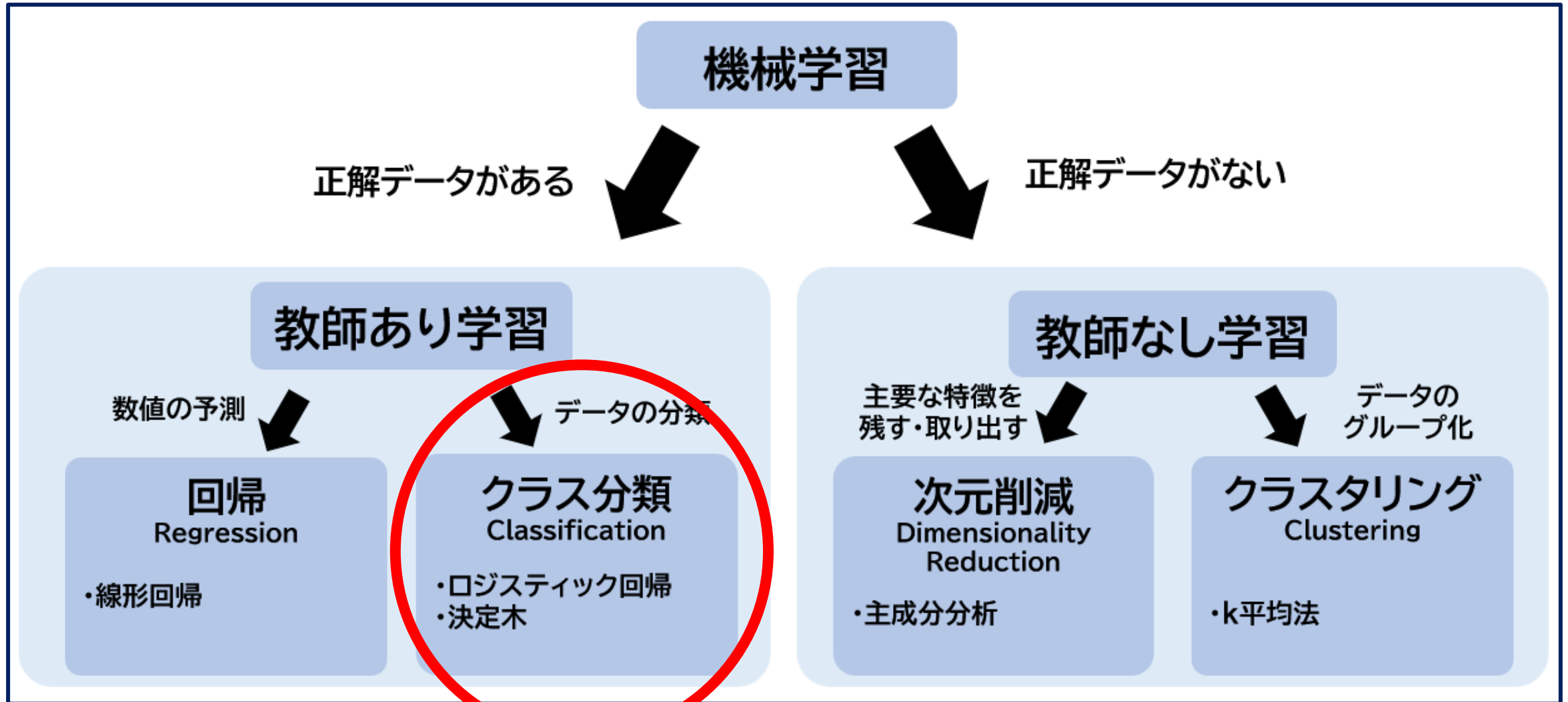
Theory12.機械学習と人工知能

- 計算機の処理能力の向上や機械学習、人工知能で用いられる手法の目覚ましい発展により、多様かつ多量の情報が適切に処理できるようになった。
- 機械学習には教師あり学習と教師なし学習がある。



①人工知能入門 モデルを用いた画像認識

<https://www.youtube.com/watch?v=EQAmucowTsc>



①モデルを用いた画像認識 自動で顔にぼかしを入れよう！

生徒の実験・感想・考察

- aiで生成した画像の**キャラクターの目は認識された**のが面白かった。バーチャルアイドルなどのイラスト系はぜんぶダメでした。
 - syamuのように**サングラスを掛けていると反応しません**でした→目のパーツを認識できないから？
 - 月面にある顔に見えるものはAIには認識されなかった。
 - 人ではできたかロボットでは顔として読み込むことができなかった
 - 人の感情をaiが持てるのかも試してみたいです。
- モナリザなどの人の絵などは本物の絵も簡単に描いたイラストでも顔として認識された。しかし、ドラえもんや猫などは認識されなかった。
 - ほかの人の画像認識の結果を聞いてもイラスト系の顔認識は判別しにくいことが分かった。
 - **スカートの影を目として認識**していた。他にも髪の毛の分け目が目と認識された。

データサイエンスだけ抜き出すと……

2h

人工知能入門 モデルを用いた画像認識
人工知能(AI)とデータの関係性って何？

情報II

モデルを用いた
画像認識

武善 紀之



1h

クラスタリングで似ている友達を探せ
仕組みの概念的な理解を試してみる

情報II

クラスタリング

武善 紀之



3h

主成分分析で「好きなもののマップ」を作ろう
実際にデータサイエンスを使ってみる！

情報II

主成分分析による
次元縮約

武善 紀之



②趣味を用いたクラスタリング

機械学習

正解データがある

正解データがない

教師あり学習

数値の予測

回帰

Regression

- ・線形回帰

データの分類

クラス分類

Classification

- ・ロジスティック回帰
- ・決定木

教師なし学習

主要な特徴を
残す・取り出す

次元削減

Dimensionality
Reduction

- ・主成分分析

データの
グループ化

クラスタリング

Clustering

- ・k平均法

②趣味を用いたクラスタリング

2番目にやる目的

数学的にイメージしやすい

30分程度でさくっと終わる！

<https://www.youtube.com/watch?v=g2qkgKglA8c>

クラスみんな、気が合う人は誰だろう？

アンケートをやってみよう！

どれくらい好きか、選んで下さい。

5(とても好き) / 4(好き) / 3(普通)
2(あまり好きではない) / 1(好きではない)

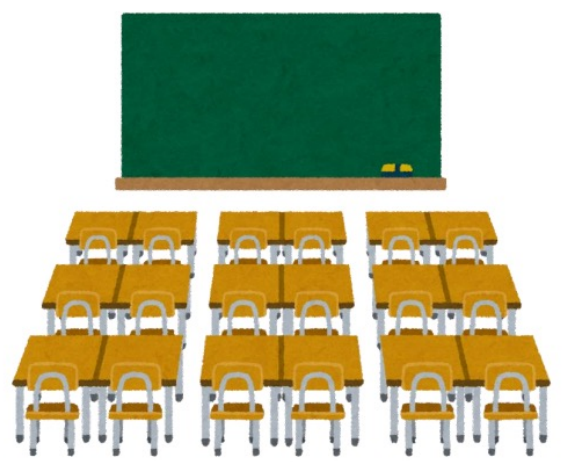
Q1. プログラミング 5 4 3 2 1

Q2. 野球 5 4 3 2 1

Q3. 楽器演奏 5 4 3 2 1

⋮

⋮



	洋楽	ドラマ	野球	JPOP	プログラミライブ	バスケ	カラオケ	ペンギン	水泳	登山	舞台鑑賞	パン	ダンス	料理	楽器演奏	屋敷	動画鑑賞	犬	猫		
生徒1		3	5	5	2	1	4	1	2	3	3	3	2	3	4	5	1	4	3	5	5
生徒2	3	3	3	2	3	1	5	1	2	5	3	5	5	4	2	1	2	2	3	5	
生徒3	1	5	4	3	2	3	4	4	1	4	2	4	5	3	3	3	4	5	4	1	
生徒4	4	3	5	2	1	3	5	4	1	2	2	2	4	5	3	2	4	4	1	2	
生徒5	2	4	2	3	4	1	3	4	3	4	5	4	3	1	1	2	5	2	3	2	
生徒6	3	5	1	2	5	2	5	5	2	1	3	5	2	2	2	5	3	3	4	4	
生徒7	4	2	1	1	2	5	5	4	3	3	5	2	1	5	1	3	2	1	2	3	
生徒8	2	3	3	3	1	3	1	3	4	3	4	5	5	5	4	1	4	1	4	3	
生徒9	3	2	4	1	2	2	5	2	4	2	3	2	4	4	5	4	2	4	1	4	
生徒10	4	5	4	2	1	2	4	3	3	4	4	4	3	5	5	4	3	4	5	1	
生徒11	5	5	1	5	2	3	1	5	3	2	2	3	5	5	3	5	5	2	5	5	
生徒12	5	1	1	5	5	4	1	5	4	2	1	4	3	5	1	5	4	4	2	1	
生徒13	5	1	1	3	5	2	1	5	2	4	3	1	4	2	5	4	1	3	4	2	
生徒14	4	2	4	1	3	1	2	3	4	2	3	5	3	3	5	3	2	1	1	1	
生徒15	2	2	2	1	2	3	4	4	4	5	4	2	1	2	4	2	4	5	2	3	
生徒16	5	1	4	4	2	3	4	4	3	4	3	1	4	1	4	3	3	2	5	2	
生徒17	5	2	4	1	2	1	3	2	4	3	3	2	1	3	4	4	4	4	4	5	
生徒18	2	5	2	3	1	3	5	5	1	1	3	3	2	5	3	4	2	3	4	2	
生徒19	1	1	3	3	1	5	5	4	2	4	2	1	2	4	4	3	5	5	4	1	
生徒20	1	2	2	1	5	2	1	4	2	3	2	1	5	5	5	3	1	3	2	3	
生徒21	3	2	4	4	2	1	1	4	2	5	2	1	1	3	3	2	1	5	3	1	
生徒22	2	3	5	4	2	2	2	2	2	5	4	5	2	2	4	1	2	4	4	4	
生徒23	5	2	2	1	1	1	5	4	5	2	2	1	3	1	5	5	1	3	4	5	
生徒24	4	5	2	5	1	4	3	5	4	3	2	3	1	5	5	5	2	5	4	3	
生徒25	5	1	4	4	1	1	1	2	1	5	4	5	3	4	5	2	2	5	3	2	
生徒26	5	5	2	3	1	2	5	5	1	3	4	3	2	3	2	2	5	2	4	2	
生徒27	3	2	4	5	2	4	1	4	1	3	1	4	4	5	2	4	5	5	2	3	
生徒28	5	3	2	5	4	2	3	3	2	1	3	5	1	2	5	4	3	5	3	4	

どうやって似ている人を探す？

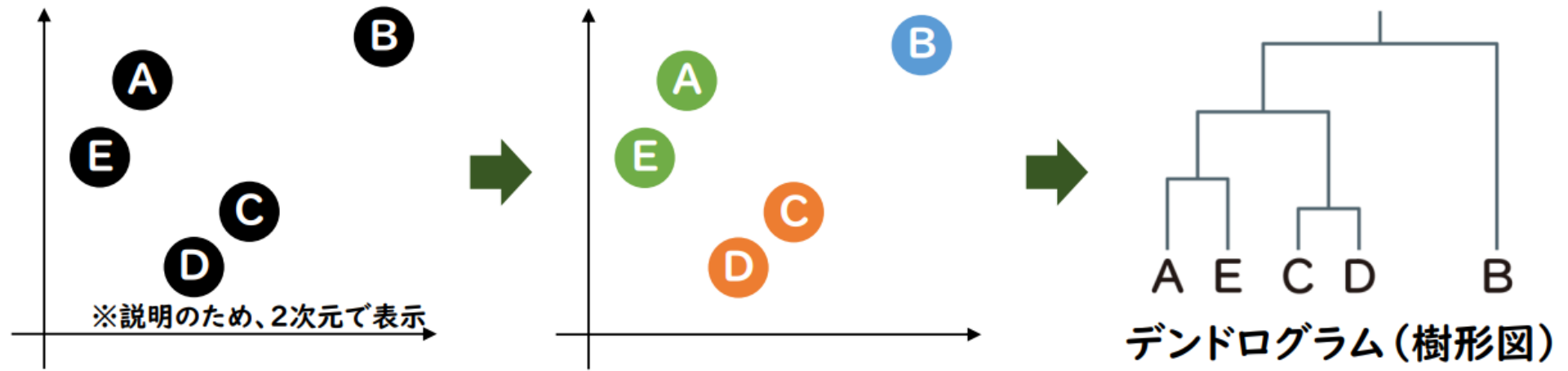
②趣味を用いたクラスタリング

クラスタリングの説明

<https://www.youtube.com/watch?v=g2qkgKglA8c>

クラスタリング

データ間の距離を計算して、近いもの同士を結びつけて行く。



	洋楽	ドラマ	野球	JPOP	プログラミライブ	バスケ	カラオケ	ペンギン	水泳	登山	舞台鑑賞	パン	ダンス
生徒1	3	5	5	2	1	4	1	2	3	3	3	2	3
生徒2	3	3	3	2	3	1	5	1	2	5	3	5	5
生徒3	1	5	4	3	2	3	4	4	1	4	2	4	5
生徒4	4	3	5	2	1	3	5	4	1	2	2	2	4
生徒5	2	4	2	3	4	1	3	4	3	4	5	4	3
生徒6	3	5	1	2	5	2	5	5	2	1	3	5	2
生徒7	4	2	1	1	2	5	5	4	3	3	5	2	1
生徒8	2	3	3	3	1	3	1	3	4	3	4	5	5
生徒9	3	2	4	1	2	2	5	2	4	2	3	2	4
生徒10	4	5	4	2	1	2	4	3	3	4	4	4	3
生徒11	5	5	1	5	2	3	1	5	3	2	2	3	5
生徒12	5	1	1	5	5	4	1	5	4	2	1	4	3
生徒13	5	1	1	3	5	2	1	5	2	4	3	1	4
生徒14	4	2	4	1	3	1	2	3	4	2	3	5	3
生徒15	2	2	2	1	2	3	4	4	4	5	4	2	1

3次元以上のデータにも対応!

②趣味を用いたクラスタリング

<https://www.youtube.com/watch?v=g2qkgKglA8c>

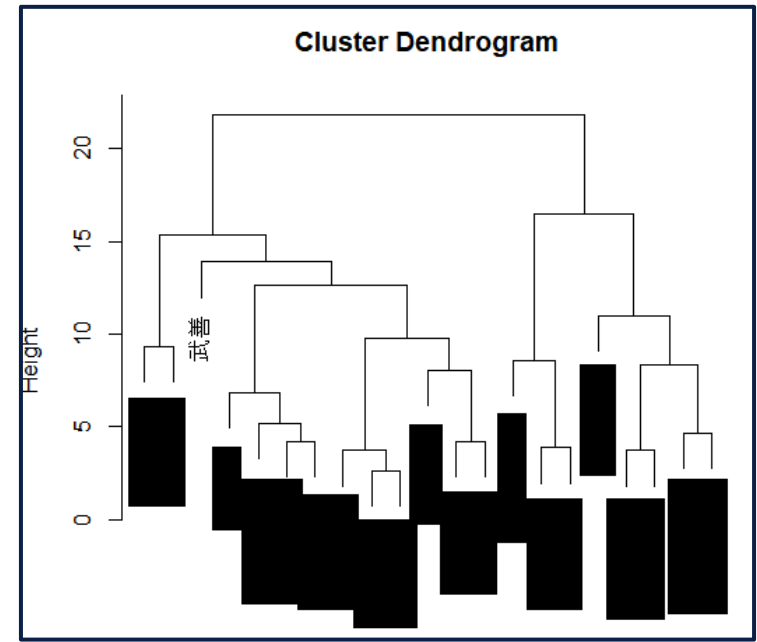
- (1) 共同編集スプレッドシートに1人1趣味入力
- (2) 1~5の5段階で「好き度」を記入
- (3) 記入されたシートを配信→Rでクラスタリング



自分の名前の	ペンギン	ゲーム	サッカー	睡眠	youtube	数学	水泳	野球	デイズニー	チョコ	本	カービィ	バス	スケ	ラーメン	散歩	運動
武善 紀之	5	5	2	5	3	4	4	2	3	4	5	4	2	2	4	5	3
	2	5	2	3	3	2	2	4	4	4	3	4	2	1	4	3	3
	4	5	5	4	5	2	4	2	3	4	2	4	1	2	5	3	4
	4	3	2	5	4	3	3	2	2	3	2	2	5	3	4	5	3
	4	4	5	5	5	2	3	2	4	5	2	5	3	3	5	5	4
	5	5	3	5	4	3	4	5	3	5	2	2	2	2	5	3	3
	2	4	3	5	5	2	5	3	5	3	3	3	2	3	4	3	3
	3	3	3	4	4	5	2	5	4	2	3	2	3	2	2	2	4
	2	4	1	5	5	3	4	2	5	5	4	2	2	2	5	5	2
	5	5	3	5	3	5	1	1	3	5	5	4	1	2	1	1	1
	2	5	2	5	5	5	1	1	3	4	5	1	2	1	5	2	1
	2	5	3	4	4	3	3	5	3	4	2	5	1	2	4	3	4
	1	5	2	5	5	4	5	5	3	2	1	3	5	4	5	4	5
	5	5	1	5	5	1	1	1	5	5	1	1	1	5	5	5	5
	3	4	2	5	5	3	3	4	3	5	2	2	2	2	5	5	4

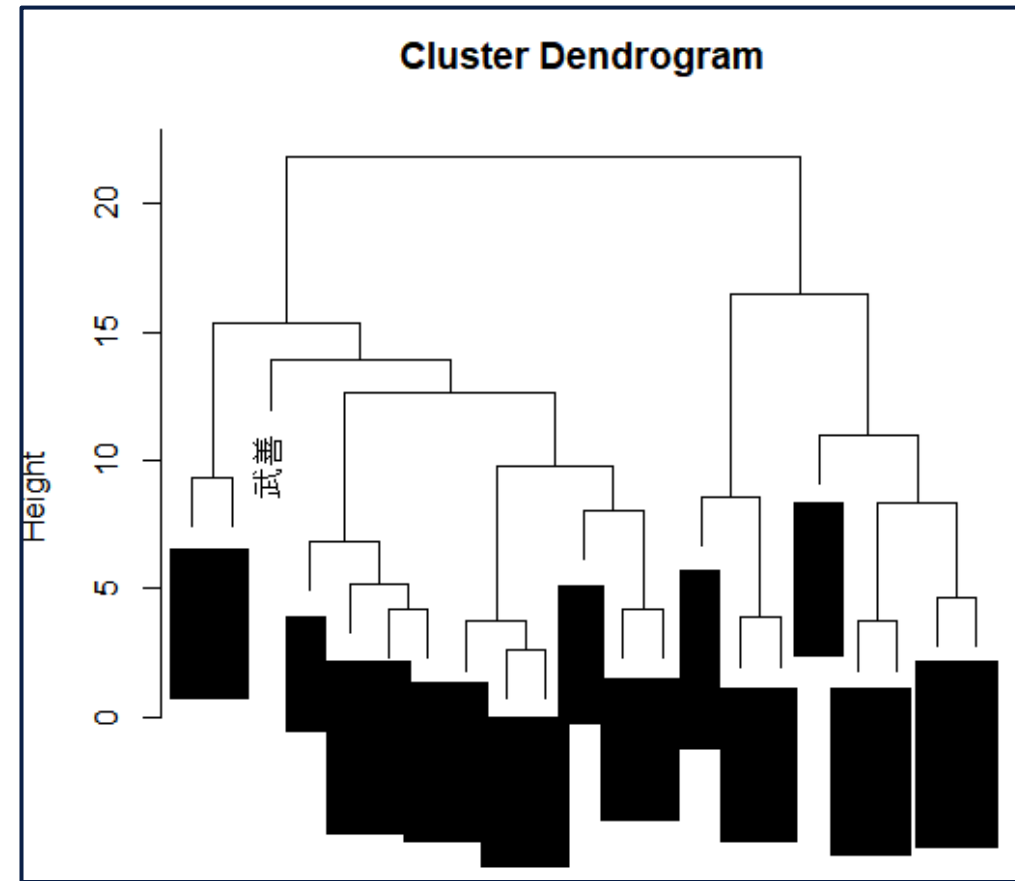


```
syumi <- read.csv("syumi.csv",h=T,row.names=1)
syumi.d <- dist(syumi)
syumi.c <- hclust(syumi.d)
plot(syumi.c)
```



②趣味を用いたクラスタリング

<https://www.youtube.com/watch?v=g2qkgKglA8c>



生徒から本当に出てきた言葉

「でも、これって何に基づいているの…？」

データサイエンスだけ抜き出すと……

2h

人工知能入門 モデルを用いた画像認識
人工知能(AI)とデータの関係性って何？

情報II

モデルを用いた
画像認識

武善 紀之



1h

クラスタリングで似ている友達を探せ
仕組みの概念的な理解を試してみる

情報II

クラスタリング

武善 紀之



3h

主成分分析で「好きなもののマップ」を作ろう
実際にデータサイエンスを使ってみる！

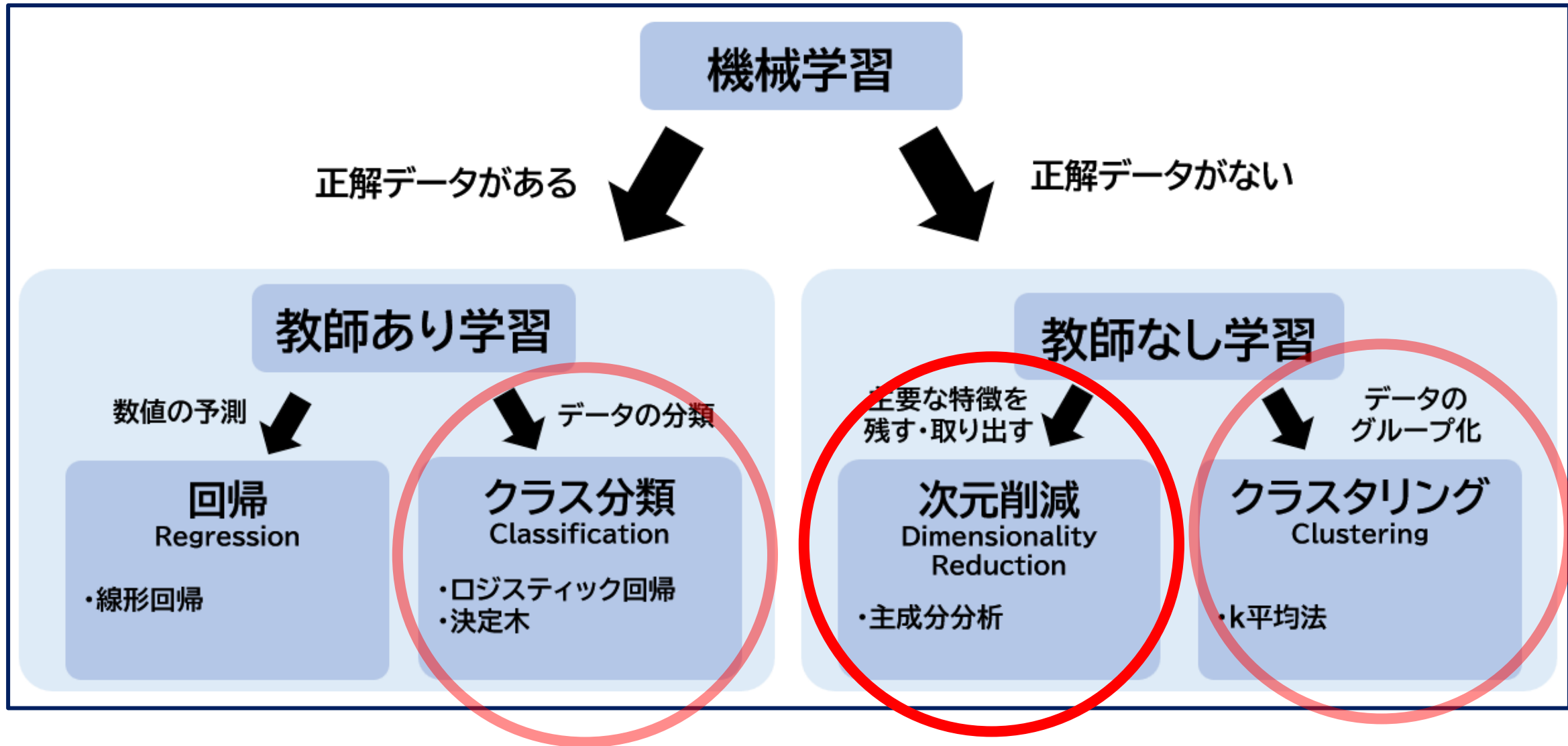
情報II

主成分分析による
次元縮約

武善 紀之



③自分の好きなものをマップ化→主成分分析



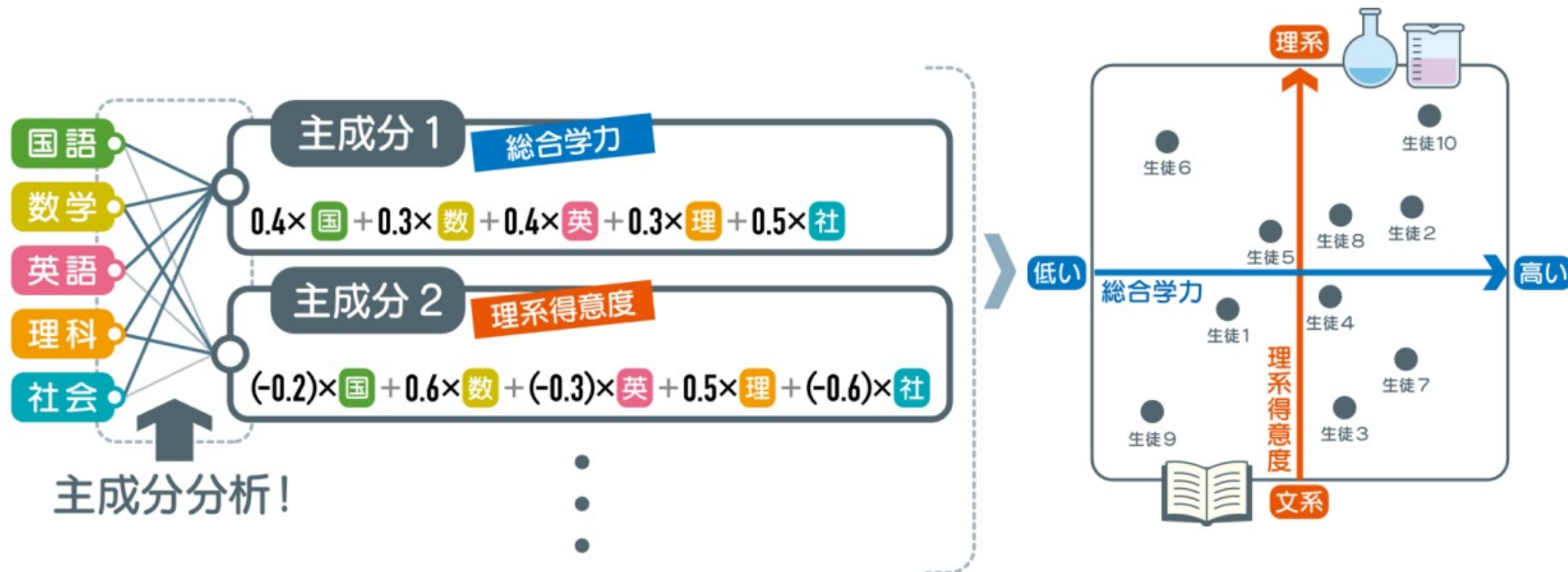
③自分の好きなものをマップ化→主成分分析

3番目にやる目的 いよいよ本格的に！

<https://www.youtube.com/watch?v=rKwjOnzXw8U>

主成分分析

多次元のデータを低次元の主成分に圧縮（要約）することで、データの構造を可視化できる。



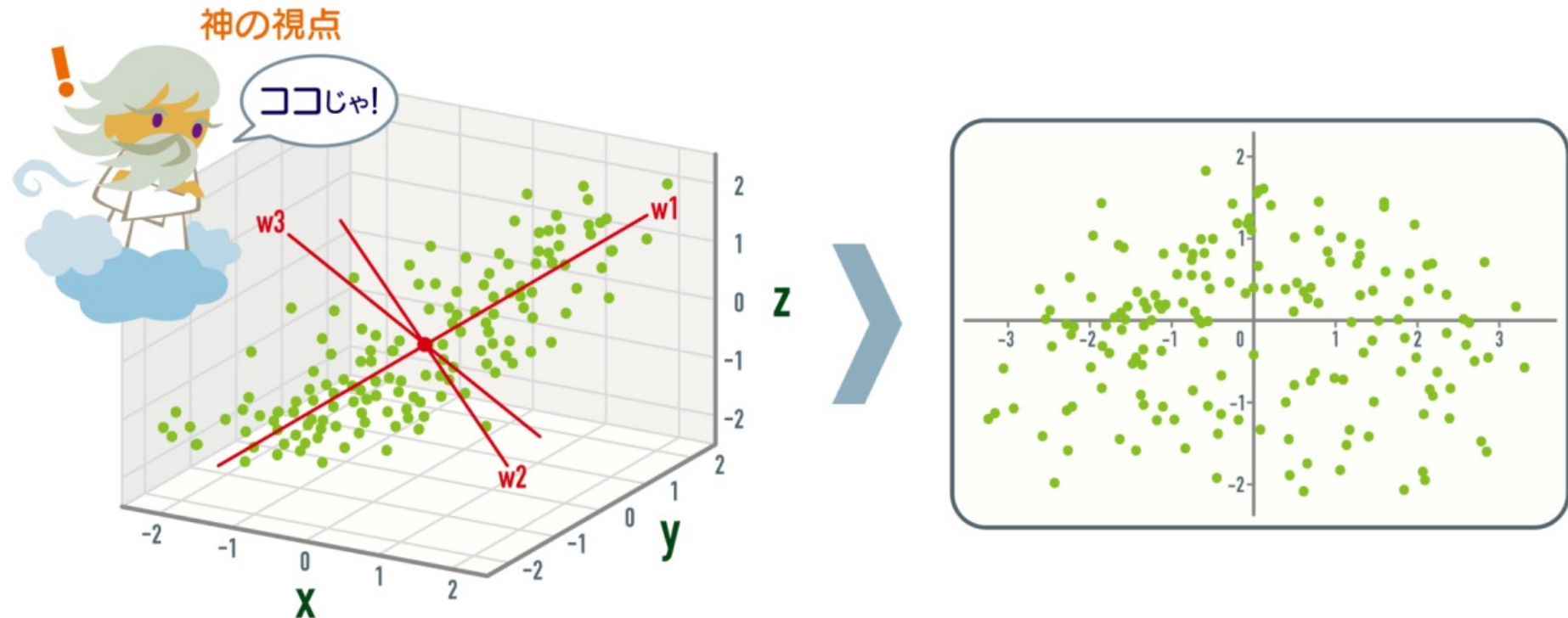
③自分の好きなものをマップ化→主成分分析

仕組み

<https://www.youtube.com/watch?v=rKwjOnzXw8U>

圧縮(要約)のイメージ

散らばりを最大化するように、空間上で軸を決める。
(一番散らばって見えるところで、見下ろすイメージ)



③自分の好きなものをマップ化→主成分分析

• 3時間構成

- 1時間目:同一事例で実演
(ヤマザキパンの栄養成分表)

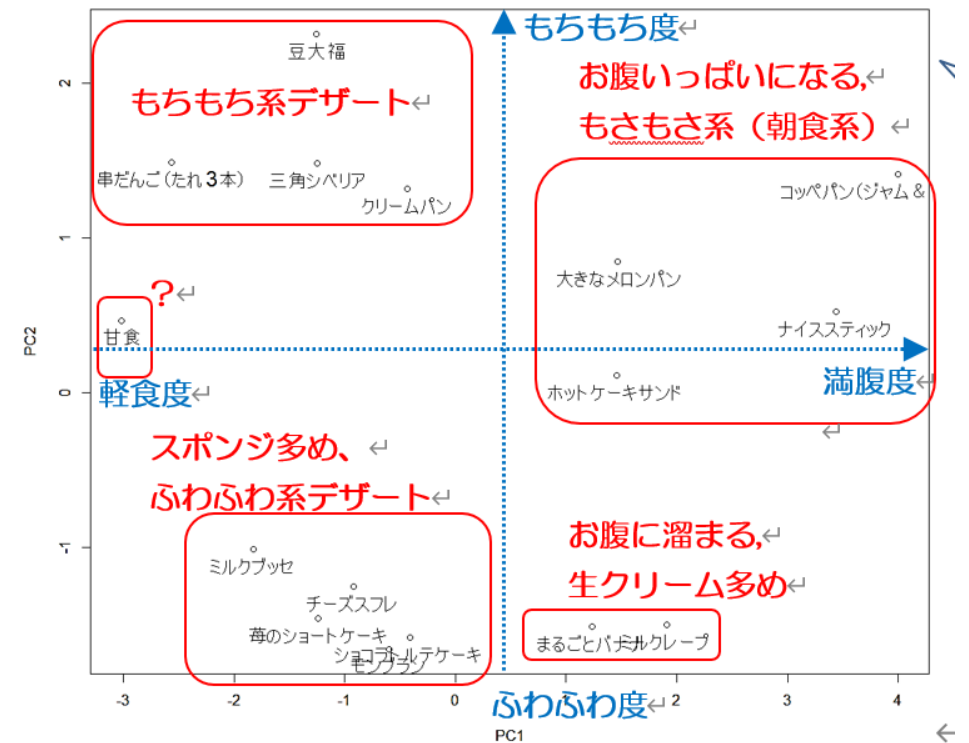
	A	B	C	D	E	F	G	H	I	J
1	熱量(kcal)	たんぱく質(g)	脂質(g)	炭水化物(g)	食塩相当量	飽和脂肪酸	トランス脂	コレステロール	(mg)	
2	苺のショー	237	2.6	15.4	22	0.2	9.6	0	46	
3	まるごとパ	447	5.3	24.2	52.1	0.3	11.1	0	93	
4	モンブラン	301	4.1	15.9	35.3	0.2	6.8	0	117	
5	三角シベリ	294	4.6	3.6	60.9	0.2	0.4	0	31	
6	ミルクレー	373	7.9	24.7	29.9	0.4	11	0.3	82	
7	ショコラト	302	3.4	20.7	25.6	0.2	12	0	30	
8	チーズスフ	224	4.9	11.3	25.9	0.4	5.9	0	112	
9	串だんご	141	1.9	0.2	33	0.5	0.1	0	0	
10	甘食	109	2	3.3	17.9	0.3	0.4	0	17	
11	豆大福	264	4.9	0.3	60.4	0.4	0.1	0	0	
12	ミルクブッ	216	2.5	13.7	20.6	0.2	3.9	0	64	
13	コッペパン	525	9.3	25.7	64	0.9	8.5	0.4	8	
14	ナイスステ	480	7.4	28	49.7	0.9	9.1	0.4	23	
15	大きなメロ	411	9.1	12.3	66	0.6	6.4	0	69	
16	ホットケー	405	6.2	20.5	49	0.9	6.3	0	77	
17	クリームパ	283	8	7.9	44.9	0.4	1.8	0	23	



<https://www.youtube.com/watch?v=rKwjOnzXw8U>

```

youhi <- read.csv("youhi.csv",h=T,row.names=1)
result <- prcomp(youhi,scale=T)
plot(result$x)
text(result$x,rownames(youhi),pos=1,cex=1.4)
    
```



- 2~3時間目:データ収集&分析

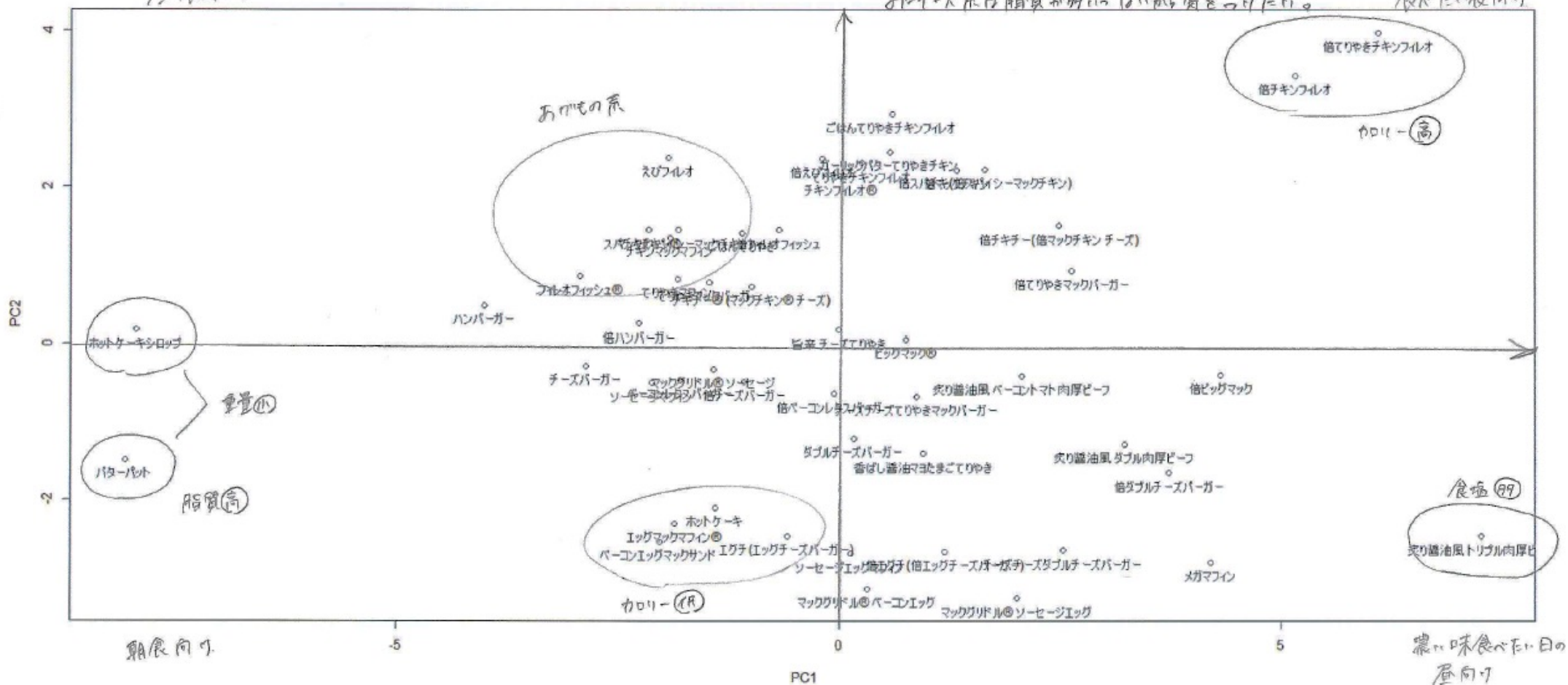
③自分の好きなものをマップ化→主成分分析

マクドナルドのメニュー

変数 製品重量・エネルギー・たんぱく質・脂質・炭水化物・ナトリウム・カリウム・カルシウム・リン・鉄・ビタミンA・ビタミンB1・ビタミンB2・ナイアシン・ビタミンC
コレステロール・食物繊維・食塩相当量

感想 主成分分析をして分析すると、普段は自分で気がつけないような共通点があるものもあると気がついた。ごはんてりやまは意外とカロリーが高くて、ごはんぱんよりも低カロリーだと思ってたが驚いた。
おチキンス系は脂質が多いのばいから気を付けてた。 食べた後夜向く

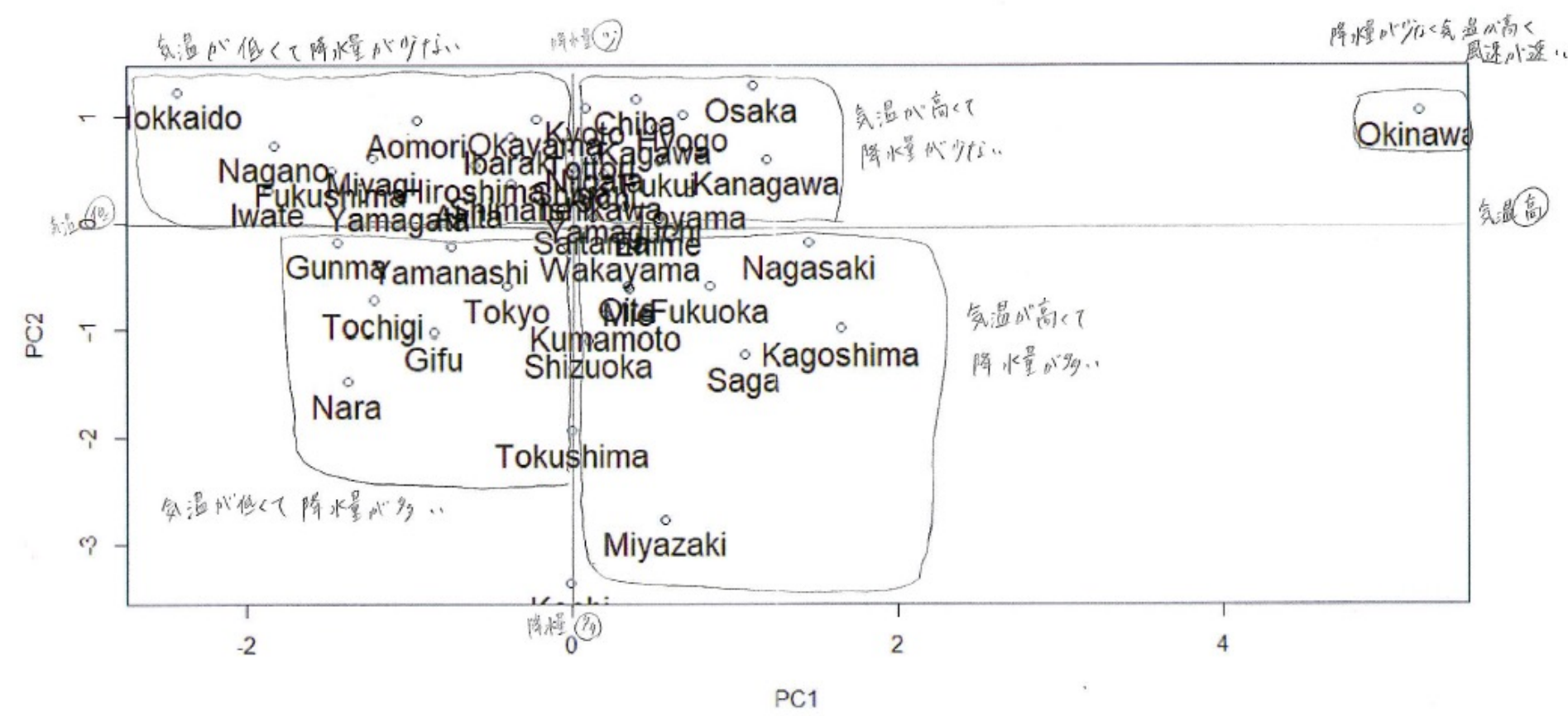
あ、わりと手軽に展開された日の昼食向



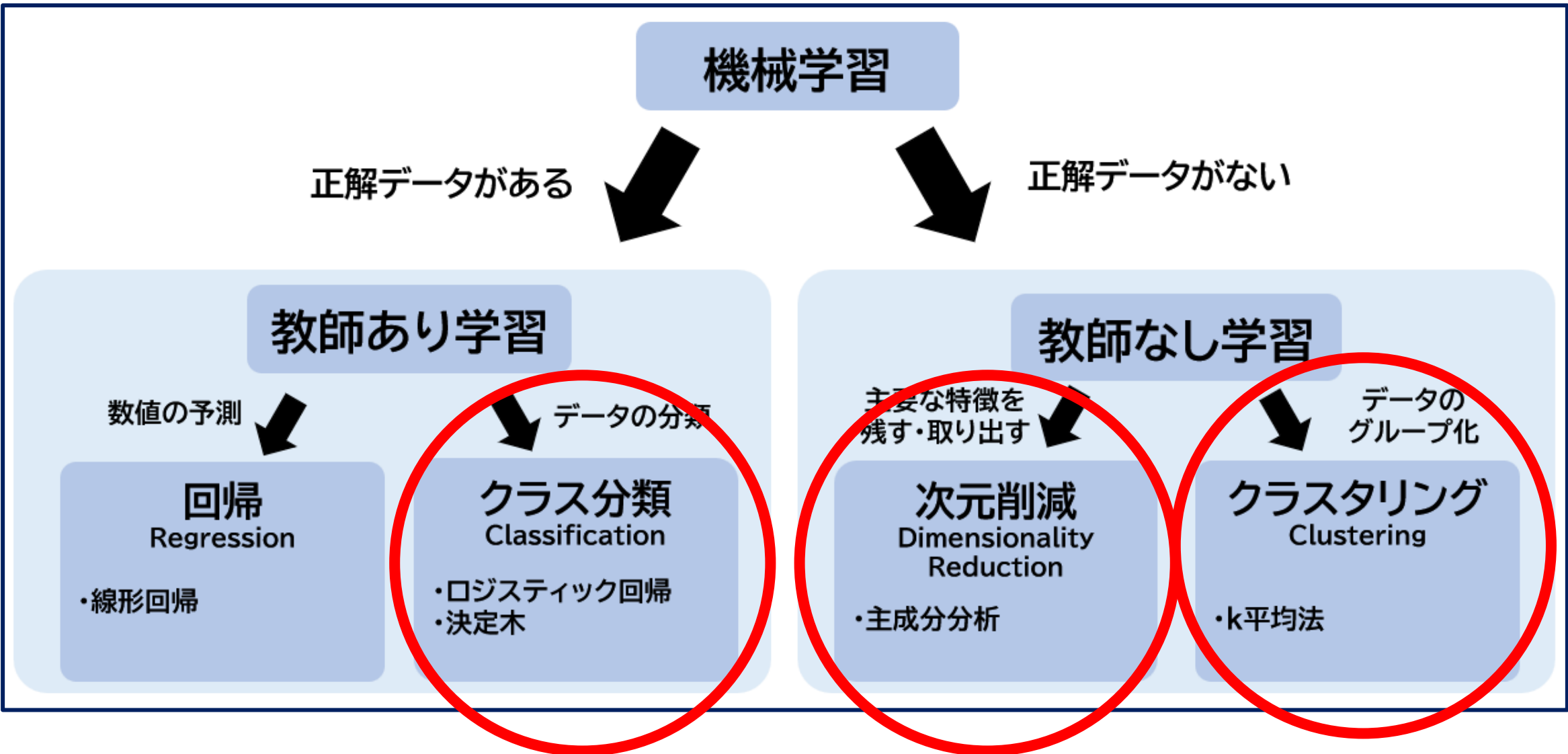
③自分の好きなものをマップ化→主成分分析

気象データ (降水量・気温・風速)

降水量、気温、風速：1980～2015年の各県の8月の平均値
沖縄は他の県と比べて気候が大きく異なる。
8月は1年間の中で台風が最も上陸する月なので、
風速も他県の平均値を大きく上回っている。
表の中心辺りに多くの県が集まっていて、似たような気候の県が多い。



以上でデータサイエンス編終了



文科省「情報Ⅱ 解説動画」データサイエンス編のすべて

情報Ⅱ 解説動画

<https://www.nttls-edu.jp/joho/>

- 重回帰分析を用いた予測「睡眠時間を他の行動時間から予測しよう」
- 主成分分析による次元縮約「データを圧縮して、関係を見よう！」
- 機械学習による分類「手書きの数字をコンピュータに認識させよう」
- クラスタリング「自分と近い性格の人は誰？」
- ニューラルネットワークによる分類「より複雑な画像をコンピュータに認識させよう」
- モデルを用いた画像認識「自動で顔にぼかしを入れよう！」

情報とデータサイエンス



情報とデータサイエンスについて



重回帰分析を用いた予測
睡眠時間を他の行動時間から予測しよう



主成分分析による次元縮約
データを圧縮して関係を見よう！



機械学習による分類
手書きの数字をコンピュータに認識させよう



クラスタリング
自分と近い性格の人は誰？



ニューラルネットワークによる分類
より複雑な画像をコンピュータに認識させよう



モデルを用いた画像認識
自動で顔にぼかしを入れよう！

この実習を終えたあとの生徒達 この構図のイメージが、出来ているはず

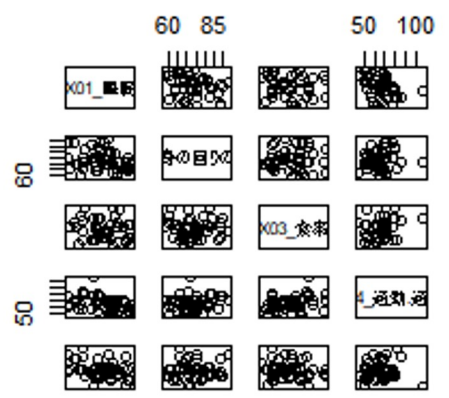
多次元のデータ

情報 I

2つずつで
頑張ってみる

相関行列

	01 睡眠	02 身の回り	03 食事	04 通勤・通学	05 仕事	06 学業
01 睡眠						
02 身の回り	-0.2308					
03 食事	-0.129	0.04927				
04 通勤・通学	-0.3563	0.19467	0.28387			
05 仕事	-0.3274	0.08675	-0.1459	0.02845		
06 学業	-0.3835	-0.1301	-0.2055	-0.0192	-0.0992	
07 家事	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732
10 買い物	0.26421	0.02212	0.0845	0.12755	-0.4664	-0.239
11 移動(通)	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163
12 テレビ	0.35748	-0.1084	0.11338	-0.0222	-0.2739	-0.0922
13 休養(く)	0.09791	-0.1746	0.04584	-0.1438	0.2955	-0.4069
14 学習(目)	-0.1175	0.37528	-0.0702	0.03141	0.02206	-0.2363
15 趣味(好)	0.12724	-0.1379	0.11829	-0.052	-0.3413	-0.2828
16 スポー	0.03883	-0.0809	-0.0632	-0.2982	0.01111	-0.1535
20 その他	0.10732	-0.0867	-0.2132	-0.3353	-0.1576	0.01509



情報 II そのまま見る!

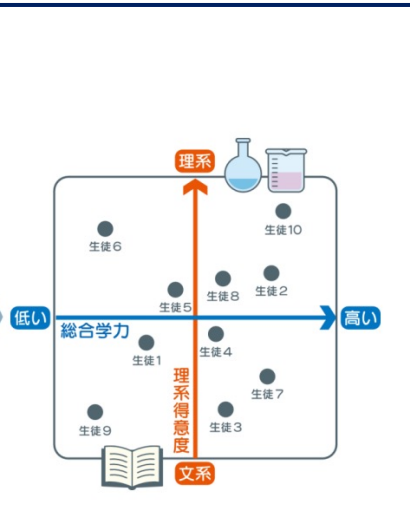
主成分分析で、次元を圧縮しよう

主成分分析!

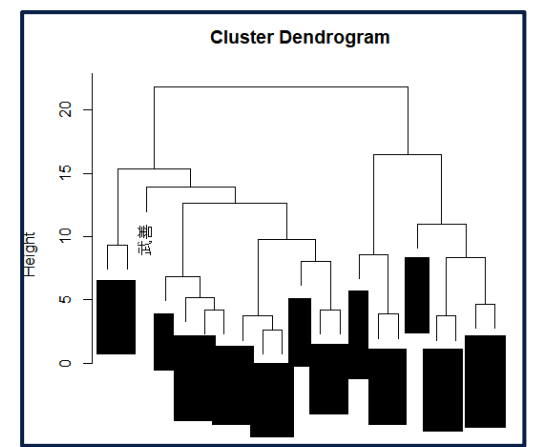
国語、数学、英語、理科、社会

主成分 1 総合学力
 $0.4 \times \text{国} + 0.3 \times \text{数} + 0.4 \times \text{英} + 0.3 \times \text{理} + 0.5 \times \text{社}$

主成分 2 理系得意度
 $(-0.2) \times \text{国} + 0.6 \times \text{数} + (-0.3) \times \text{英} + 0.5 \times \text{理} + (-0.6) \times \text{社}$



クラスタリングで、グループ化しよう



etc...

ただ、情報Ⅱ的にはこの先が大事

- 1学期は各単元における「学びの種」を蒔いた状態
- 2学期の卒業研究・卒業制作が「情報Ⅱ」のメインとなる

2023 年度『情報の科学』卒業研究

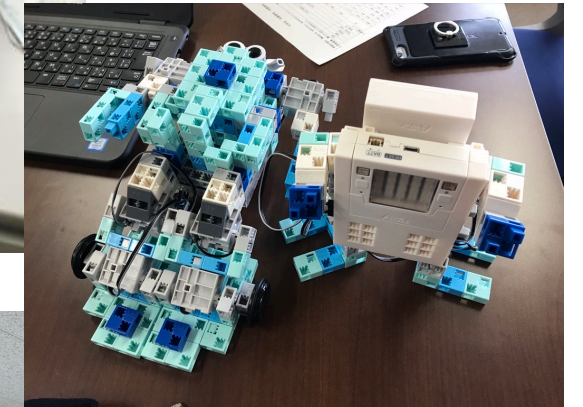
■ やること

※10/19(木)までに、Google Classroom へグループを登録する。

テーマ例一覧（複数領域に跨る場合もある）

【推奨テーマ】

番号	テーマ	備考
A1	プログラミング等による開発(ハードウェア)	micro:bit 等を使用（その他ロボット教材等あり）
A2	プログラミング等による開発(ソフトウェア)	ドリトル、Processing、表計算ソフト、Web ページなど。
B1	シミュレーション要素を含んだ調査研究	家具配置,ガチャシミュレーションの発展など
B2	データサイエンス要素を含んだ調査研究	データ分析（アンケート調査、代表値計算、グラフ化）等を行うこと
B3	情報技術・情報社会に関する調査研究	（例）化粧の工夫は、同一人物判定の結果に寄与するか？



学期末アンケート

・情報Ⅱのデータサイエンス・統計領域を学んだ感想 (「数学Ⅰ」や「情報Ⅰ」との違いを中心に)

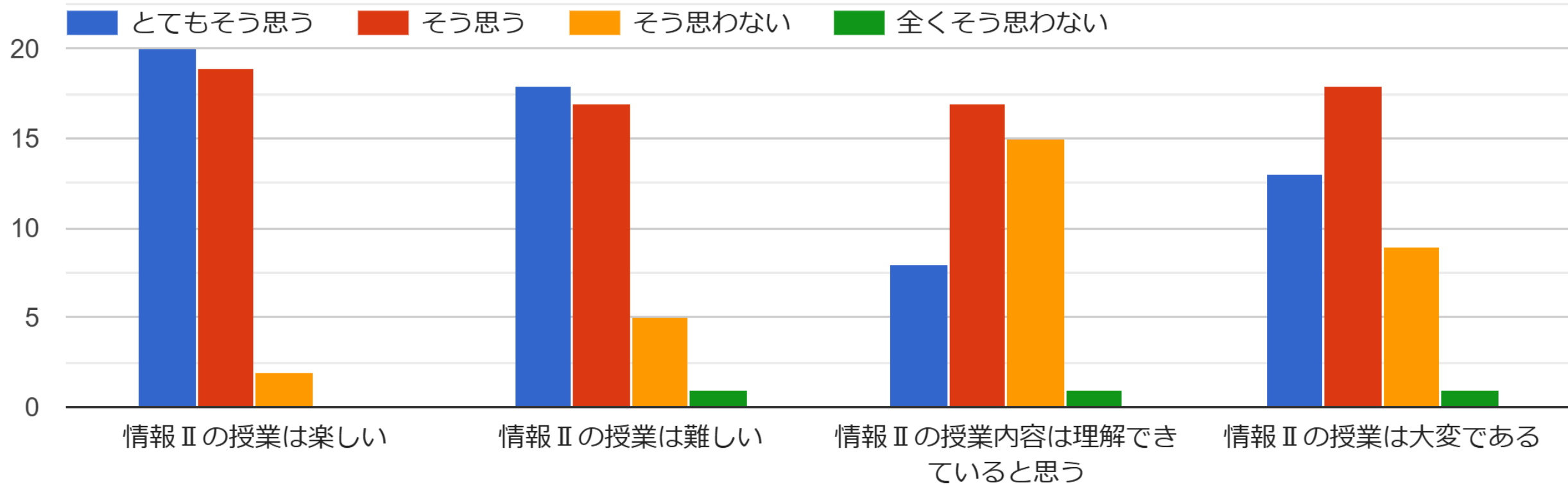
(全般的)

- ・ **たくさんの項目をひとつにまとめて傾向や偏りが一目見てわかるようになるのはすごい**と思った。
- ・ **主成分分析が人間が認識できないほど大きな次元をもとに分類できるのがすごい**と思った。
- ・ **顔認識は写真アプリなどにも応用されているのかな**と思い、**普段の生活でも使われているのかもしれない**と気づいた。
- ・ **情報Ⅰは覚えることや暗記がある程度多くてあまり得意ではありませんでしたが、情報Ⅱは実践的なものが多かった**ので楽しかったです。

(数学との比較)

- ・ **自分が数学で学んだ分散を使ってこのような主成分分析ができることを知り、とても驚きました。**
- ・ **数学においてデータは二次元でまとめるものだと思っていた**けど、主成分分析はいくつも軸があるやつをまとめられてこんな方法もあると知った
- ・ **たくさんのデータをまとめることは数学などでは難しいので、それらをいっきにまとめられるのはとても便利だ**と思った。
- ・ **正確な値ではなく、傾向を見るのが新鮮**だった。

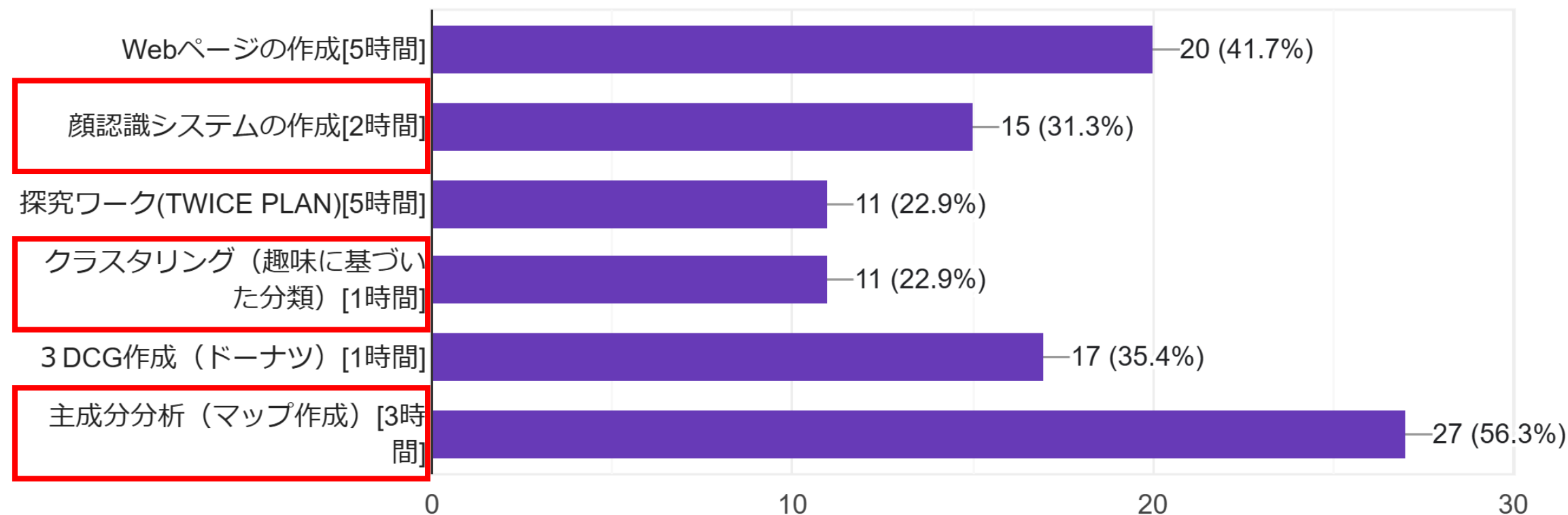
学期末アンケート(4件法)



学期末アンケート

「面白かった（楽しかった、やる気になって取り組めた）」単元（複数選択）

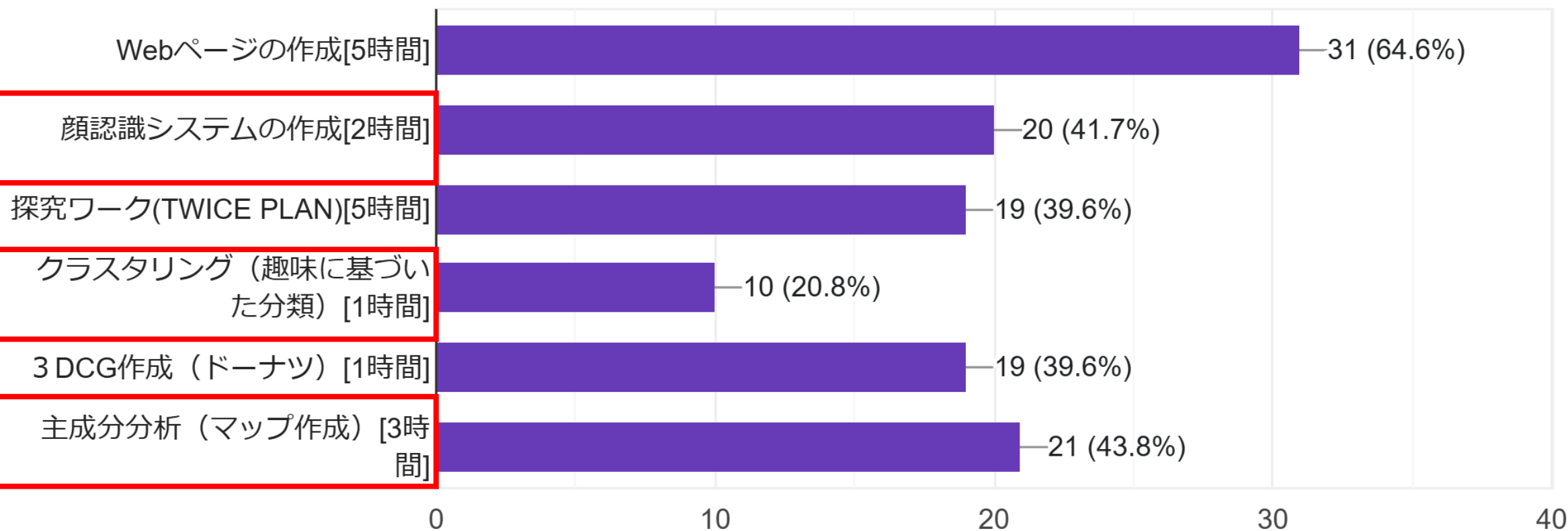
48件の回答



学期末アンケート

「学びになった」単元（複数選択）

48件の回答



この情報Ⅱに向かって、
情報Ⅰを設計する
(+入試)

Agenda.

- 1 情報Ⅰ / 情報Ⅱのデータサイエンス(学習指導要領より)
- 2 情報Ⅱにおける実践
- 3 情報Ⅰにおける実践**
- 4 データサイエンスを扱う価値3つ(個人的に)

情報Ⅰのデータサイエンス①

- 昔は色々やっていたが最近はともかく簡素に、最小限に。

仮説検定を高校生に教える

-情報科と数学科では何が違う？-



“怪しげな情報”の「作成・法則・検証」で
養うメディアリテラシー

「噂」と「広告」を題材に

- 「情報Ⅰ全体のボリュームUP」(社会と情報+情報の科学)
- 「共通テスト対策への意識」(表計算の経験は必要)
- 「総合的な探究の時間」の実施(アンケート実習は総合で)
- 3変数以上は「情報Ⅱ」、2変数の処理を「情報Ⅰ」と分離

共通テスト「情報Ⅰ」試作問題

大学入試センター 試作問題「情報Ⅰ」より

第4問 次の文章を読み、後の問い(問1～5)に答えよ。(配点 25)

次の表1は、国が実施した生活時間の実態に関する統計調査をもとに、15歳以上19歳以下の若年層について、都道府県別に平日1日の中で各生活行動に費やした時間(分)の平均値を、スマートフォン・パソコンなどの使用時間をもとにグループに分けてまとめたものの一部である。ここでは、1日のスマートフォン・パソコンなどの使用時間が1時間未満の人を表1-A、3時間以上6時間未満の人を表1-Bとしている。

表1-A：スマートフォン・パソコンなどの使用時間が

1時間未満の人の生活行動時間に関する都道府県別平均値

都道府県	睡眠(分)	身の回りの用事(分)	食事(分)	通学(分)	学業(分)	趣味・娯楽(分)
北海道	439	74	79	60	465	8
青森県	411	74	73	98	480	13
茨城県	407	61	80	79	552	11
栃木県	433	76	113	50	445	57

- 多次元の表, 箱ひげ図, 散布図は必須
- 短時間で

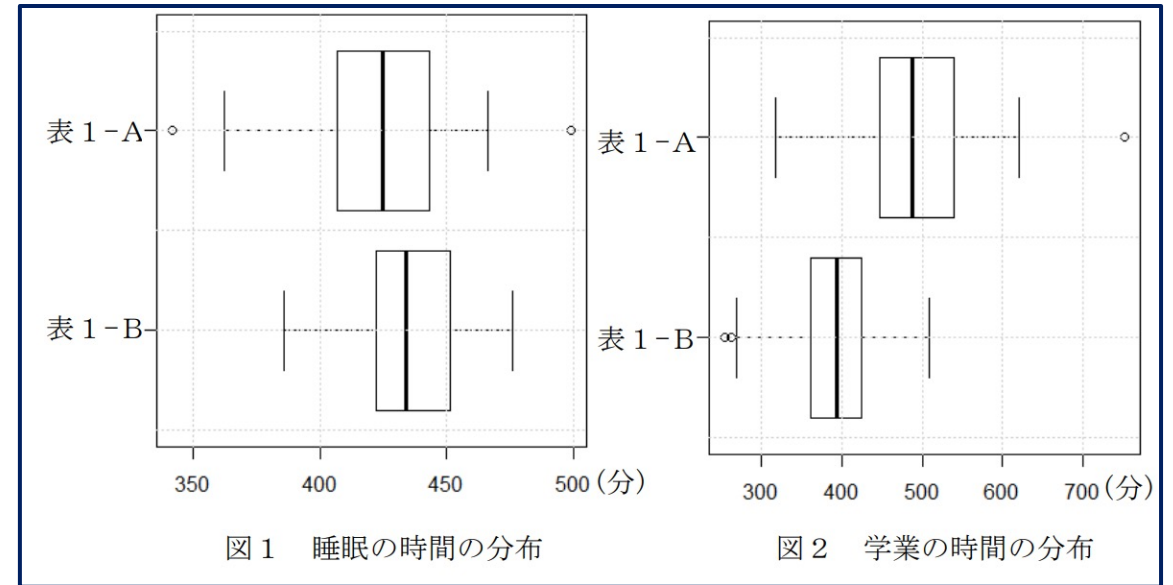
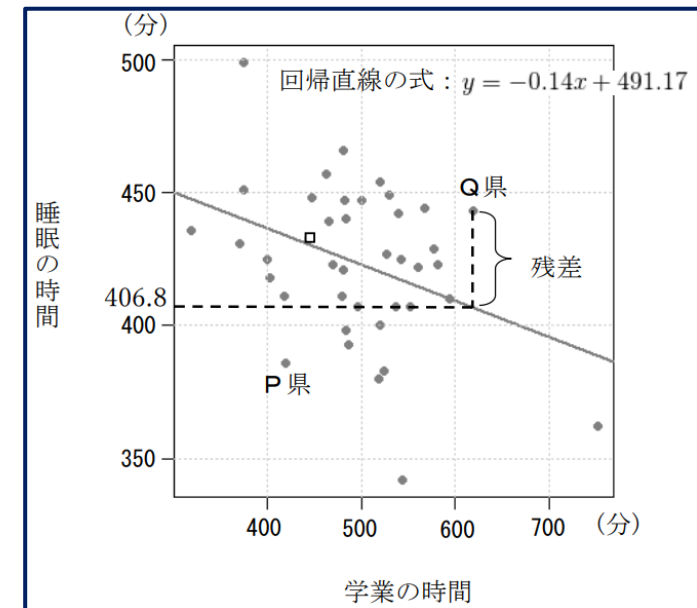


図1 睡眠の時間の分布

図2 学業の時間の分布



「これだ！」と思えた実践

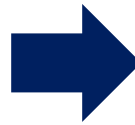
第16回全国高等学校情報教育研究会全国大会(東京大会)

「とどらん(都道府県ランキング)で教えるデータ活用の授業」

アサンプション国際中学校高等学校 岡本弘之先生

・生徒にとって身近で、興味を持ちそうなデータがたくさんある。※注意が必要
・すべて都道府県別のデータなので相関を考えるには使いやすいデータ

(<https://todo-ran.com/>)



データを整える
①北からの順に整列
②文字列を数値に変換

データを入力する
(数値を入力)

関数を入力して計算する
※相関係数 CORREL関数

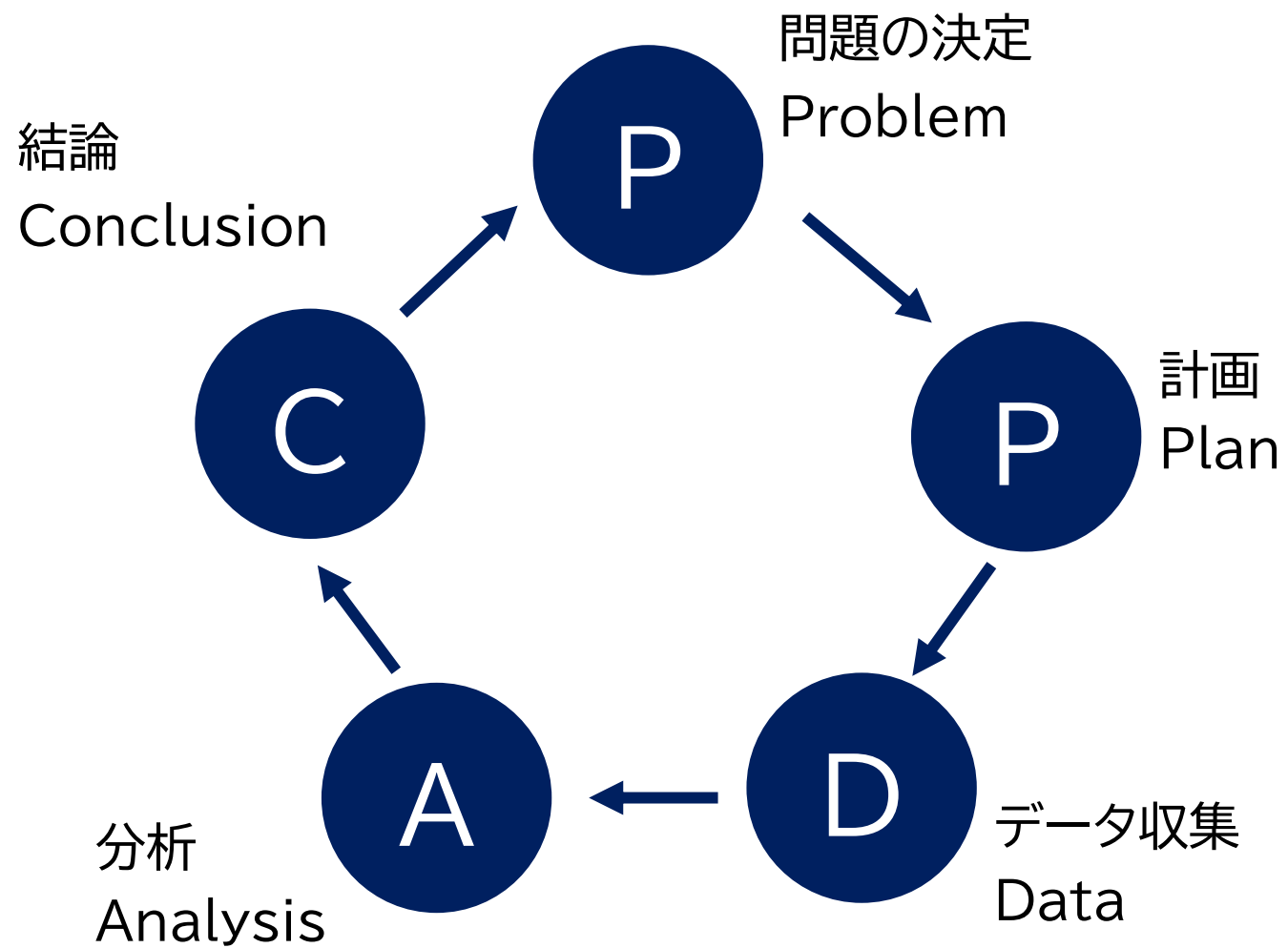
散布図も作成する

<https://todo-ran.com/>

(画像)キミの未来発見事例205

<https://www.wakuwaku-catch.net/jirei22205/>

PPDACサイクルも意識されている



3時間で実施(データサイエンス入門)

(1時間目)

・数学復習&操作習得

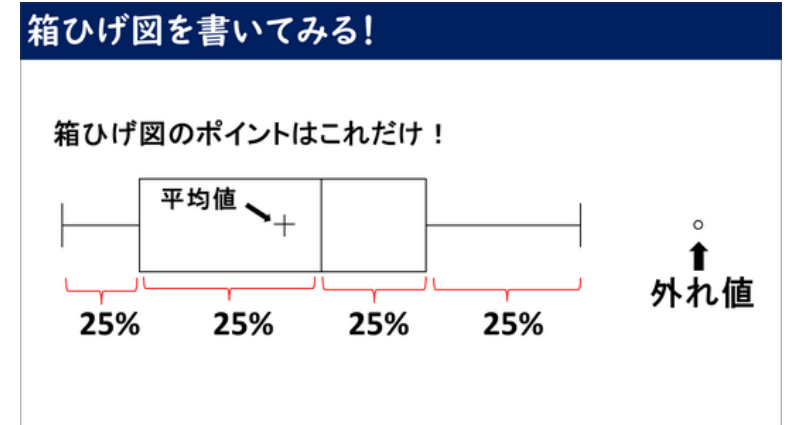
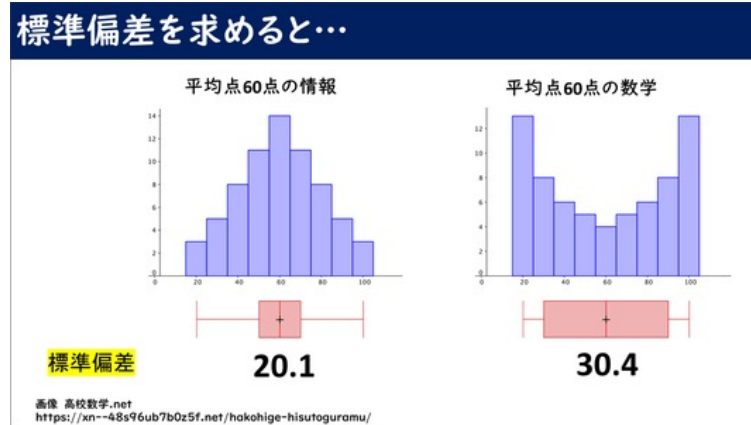
(2時間目)

・各自の分析&発表

(3時間目)

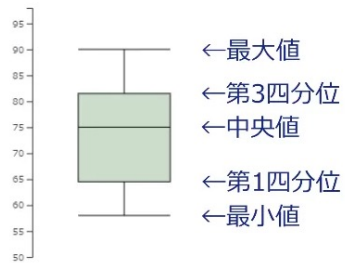
・振り返り分析

数学の復習(ヒストグラム・箱ひげ図・相関係数)



箱ヒゲ図作成サイト

統計データをあらわす時に、箱ヒゲ図を使うことがあります。箱ひげ図の書き方は、以下の図のように最大値、第3四分位、中央値、第1四分位、最小値を箱の形で示します。

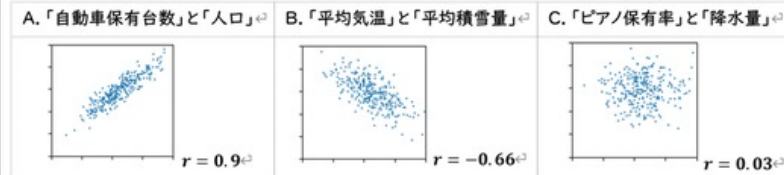


上図のように、箱があり、その両端にヒゲがあるように見えますね。

- ・箱の部分：全体の1/4から3/4まで
- ・ヒゲの部分：最小値から最大値まで

と、データの範囲が分かりやすく図示されます。

相関係数



-1.0 ~ -0.7	-0.7 ~ -0.4	-0.4 ~ -0.2	-0.2 ~ 0.2	0.2 ~ 0.4	0.4 ~ 0.7	0.7 ~ 1.0
強い負の相関	負の相関	弱い負の相関	相関なし	弱い正の相関	正の相関	強い正の相関

-1 ~ 1の値をとる。

画像: のんびり統計
https://note.com/e_doo/n/n7697b32f1a12

☆箱ひげ図も作成させる

<https://statistics.calculator.jp/basic/box-plot/>

3時間で実施(データサイエンス入門)

(1時間目)

- ・数学復習&操作習得

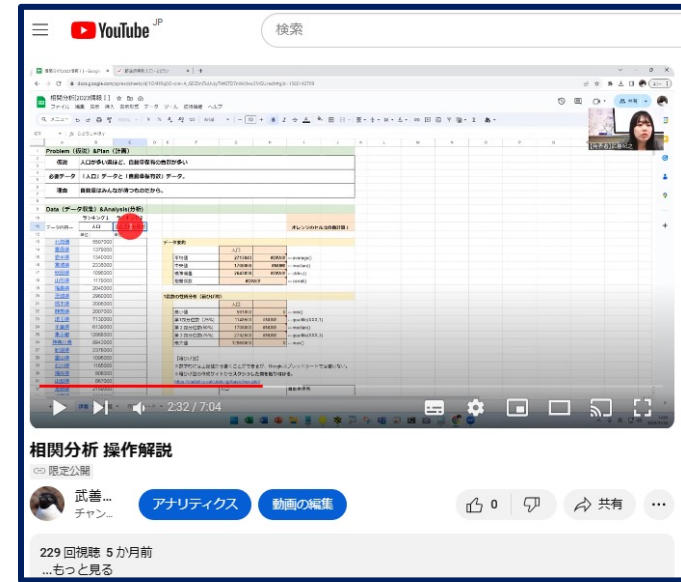
(2時間目)

- ・各自の分析&発表

(3時間目)

- ・振り返り分析

YouTubeで操作解説動画



共同編集スプレッドシートで知見共有

氏名	仮説	データ1	データ2	相関係数	次に検証したい仮説
武善 紀之	人口が多いほど、自動車を保有	人口	自動車保有台数	0.95	ガソリンスタンドの数は面積と人口、どちらに関係するた
	高齢者が多いほど、パチンコ店	人口	パチンコ店舗数	0.985	人口とパチンコ店舗数の関係
	都会になるにつれて家賃が高く	人口	家賃	-0.098	サッカーと野球の人口
	睡眠時間が短いほど年収が高い	平均年収	平均睡眠時間	-0.735	大企業数が多いと年収が高い
	降雪量が多いとスキーやスノー	競技人口	降雪量	-0.003	降雪量とスキーやスノーボードの競技人口競技人口の県別
	生活保護と人口の比例	生活保護世帯	人口	0.918	テニスと卓球の人口
	魅力度が高いほどふるさと納税	ふるさと納税額	魅力度	0.432	漁獲量とふるさと納税
	人口が多い都道府県のほうがセ	セブンの数	人口	0.967	セブンの数とファミマの数
	お菓子の消費量が多いほどお菓	お菓子の消費量	お菓子の消費量	0.705	お菓子の消費量とお菓子の消費量

3時間で実施(データサイエンス入門)

(1時間目)

- 数学復習 & 操作習得

(2時間目)

- 各自の分析 & 発表

(3時間目)

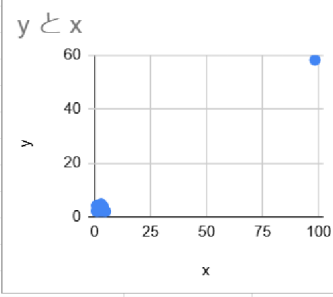
- 振り返り分析

検討1. 相関係数は本当に正しい?

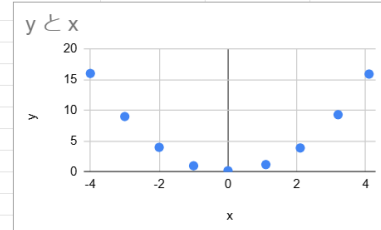
• 相関係数 0.98

すごい関係性あるな!

【データA】	
相関係数	0.996
x	y
1	2
2	2.3
3	4.8
4	3.8
5	2
3.5	2.5
1	4.2
3	0.1
1	2.4
98.4	58.2



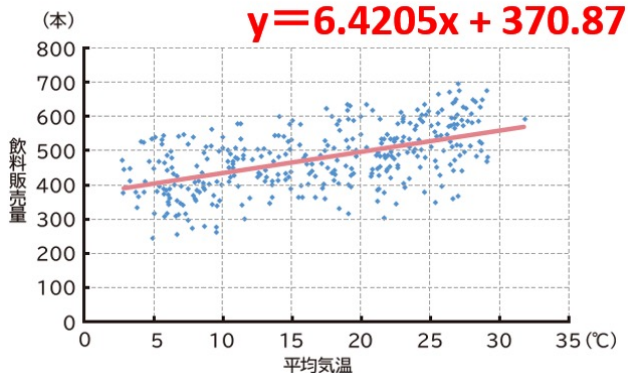
【データB】	
相関係数	0.008
x	y
-4	16
-3	9
-2	4
-1	1
0	0.2
1.1	1.2
2.1	3.9
3.2	9.3
4.1	15.9



振り返ろう2

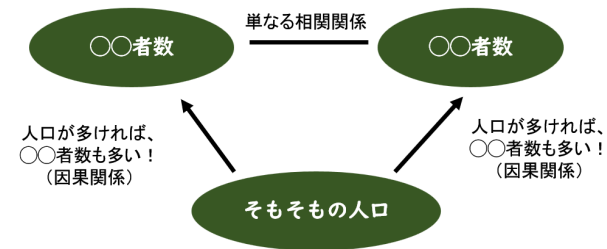
自分が調べたものは、相関関係か?
因果関係か? 因果の方向性はどちらか?
他に考えられる要因は何か?

この関係性に陥っている人が多そう。



振り返ろう3 (分析手法の発展)

単回帰分析の
数式が意味することを書いてみよう。



一応この内容で「情報Ⅰ」共通テストは網羅

大学入試センター 試作問題「情報Ⅰ」より

第4問 次の文章を読み、後の問い(問1～5)に答えよ。(配点 25)

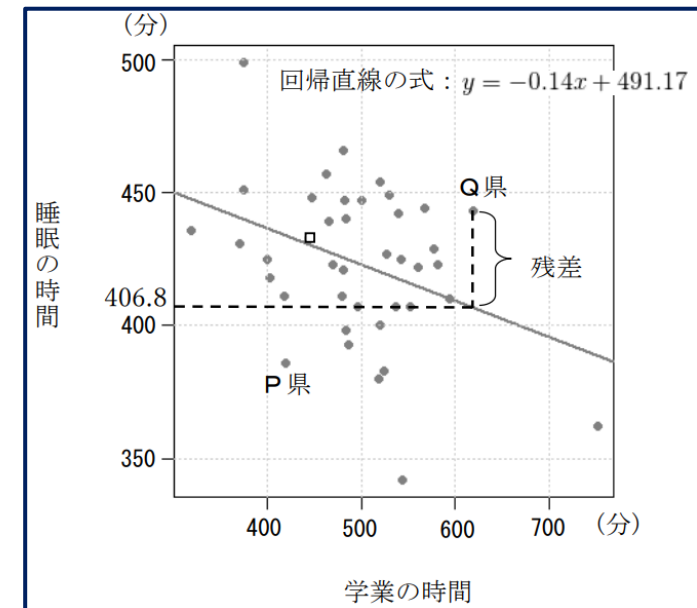
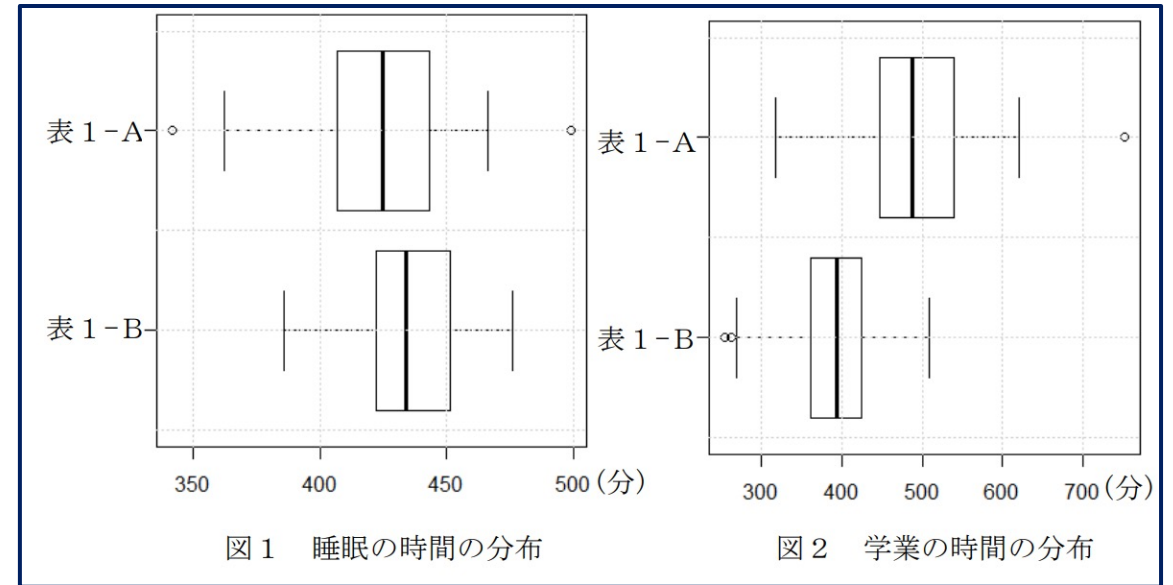
次の表1は、国が実施した生活時間の実態に関する統計調査をもとに、15歳以上19歳以下の若年層について、都道府県別に平日1日の中で各生活行動に費やした時間(分)の平均値を、スマートフォン・パソコンなどの使用時間をもとにグループに分けてまとめたものの一部である。ここでは、1日のスマートフォン・パソコンなどの使用時間が1時間未満の人を表1-A、3時間以上6時間未満の人を表1-Bとしている。

表1-A：スマートフォン・パソコンなどの使用時間が

1時間未満の人の生活行動時間に関する都道府県別平均値

都道府県	睡眠 (分)	身の回りの 用事(分)	食事 (分)	通学 (分)	学業 (分)	趣味・娯楽 (分)
北海道	439	74	79	60	465	8
青森県	411	74	73	98	480	13
茨城県	407	61	80	79	552	11
栃木県	433	76	113	50	445	57

- 多次元の表, 箱ひげ図, 散布図は必須
- 短時間で



情報Ⅰ / 情報Ⅱで培われる視点

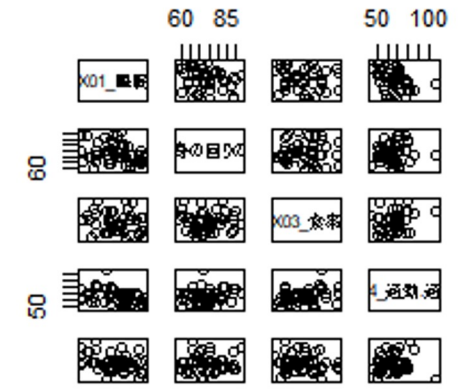
多次元のデータ

	1. 睡眠	2. 身の回り	3. 食事	4. 通勤・通学	5. 仕事	6. 学業	7. 家事	8. 買い物	9. 移動	10. テレビ	11. 休養	12. 趣味	13. その他
01. 睡眠	467	72	74	62	41	355	5	9	19	25	166	42	63
02. 身の回り	469	91	60	34	386	0	3	15	79	129	19	24	30
03. 食事	459	69	76	53	47	453	5	3	6	31	105	42	59
04. 通勤・通学	449	72	44	37	370	0	3	5	16	135	39	119	50
05. 仕事	467	61	84	45	29	379	4	6	10	19	136	32	114
06. 学業	471	74	60	52	40	377	4	5	6	40	135	39	59
07. 家事	439	60	87	64	41	439	7	2	15	29	149	36	40
08. 買い物	472	68	62	71	31	397	8	6	10	24	130	41	41
09. 移動	459	65	91	60	31	401	5	3	12	29	154	55	41
10. テレビ	457	74	78	69	41	398	7	7	12	31	125	36	67
11. 休養	459	65	81	60	31	401	5	3	12	29	154	55	41
12. 趣味	437	71	65	65	70	387	8	3	12	15	170	54	33
13. その他	451	64	66	69	33	387	5	7	11	16	124	55	36
01. 睡眠	452	69	66	77	32	399	5	8	14	39	108	33	78
02. 身の回り	475	71	74	59	12	391	7	11	6	40	140	49	62
03. 食事	475	66	75	66	68	376	3	4	10	35	125	29	64
04. 通勤・通学	469	69	78	51	48	395	3	1	16	16	135	31	79
05. 仕事	445	67	76	56	36	450	3	4	10	12	129	54	45
06. 学業	453	75	79	62	28	429	5	1	16	30	94	60	68
07. 家事	469	62	81	72	48	402	10	1	6	29	150	39	53
08. 買い物	459	73	79	69	39	418	13	6	9	45	101	36	57
09. 移動	452	62	61	62	49	390	6	3	11	21	130	47	54
10. テレビ	459	63	79	71	72	374	4	1	12	25	123	44	51
11. 休養	431	61	65	61	67	356	5	2	6	17	173	42	41
12. 趣味	462	61	63	60	47	389	5	4	6	12	144	46	47
13. その他	476	65	63	53	23	380	17	10	14	27	159	34	53
01. 睡眠	450	69	64	79	45	396	6	6	15	23	147	59	51
02. 身の回り	465	79	67	69	45	351	3	6	12	33	161	40	48
03. 食事	446	72	69	107	10	418	7	15	3	41	75	39	100
04. 通勤・通学	467	58	91	71	31	433	7	6	14	37	139	31	45
05. 仕事	464	69	72	55	55	400	9	4	12	37	139	39	27
06. 学業	449	68	75	66	47	440	1	1	9	14	152	51	37
07. 家事	424	69	72	55	55	400	9	4	12	37	139	39	27
08. 買い物	450	64	60	67	27	433	18	3	20	152	49	49	19
09. 移動	432	70	61	67	63	384	7	5	20	25	133	34	63
10. テレビ	451	78	77	72	65	401	3	3	11	31	111	41	84
11. 休養	451	78	77	72	65	401	3	3	11	31	111	41	84
12. 趣味	451	78	77	72	65	401	3	3	11	31	111	41	84
13. その他	451	78	77	72	65	401	3	3	11	31	111	41	84

情報Ⅰ

2つつで
頑張ってる

	01 睡眠	02 身の回りの	03 食事	04 通勤・通学	05 仕事	06 学業
01 睡眠						
02 身の回り	-0.2308					
03 食事	-0.129	0.04927				
04 通勤・通学	-0.3563	0.19467	0.28387			
05 仕事	-0.3274	0.08675	-0.1459	0.02845		
06 学業	-0.3835	-0.1301	-0.2055	-0.0192	-0.0992	
07 家事	0.03418	-0.2337	0.04911	0.06168	-0.1948	0.05732
10 買い物	0.26421	0.02212	0.0845	0.12755	-0.4664	-0.239
11 移動	-0.1549	0.14999	0.02546	-0.107	-0.1786	-0.2163
12 テレビ	0.35748	-0.1084	0.11338	-0.0222	-0.2739	-0.0922
13 休養	0.09791	-0.1746	0.04584	-0.1438	0.2955	-0.4069
14 学習	-0.1175	0.37528	-0.0702	0.03141	0.02206	-0.2363
15 趣味	0.12724	-0.1379	0.11829	-0.052	-0.3413	-0.2828
16 スポーツ	0.03883	-0.0809	-0.0632	-0.2982	0.01111	-0.1535
20 その他	0.10732	-0.0867	-0.2132	-0.3353	-0.1576	0.01509

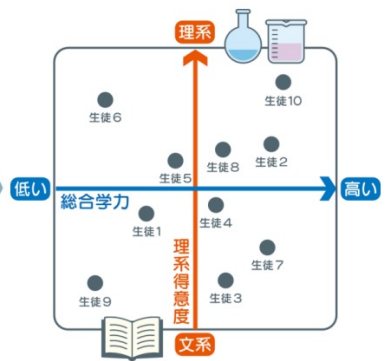
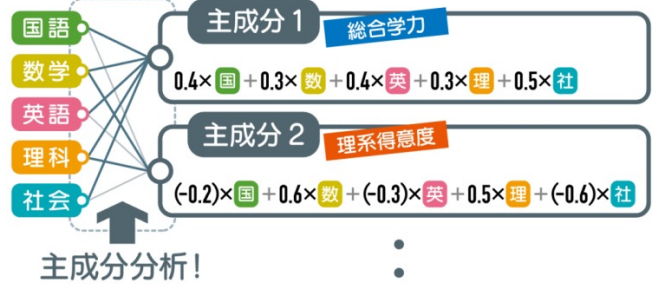


相関行列

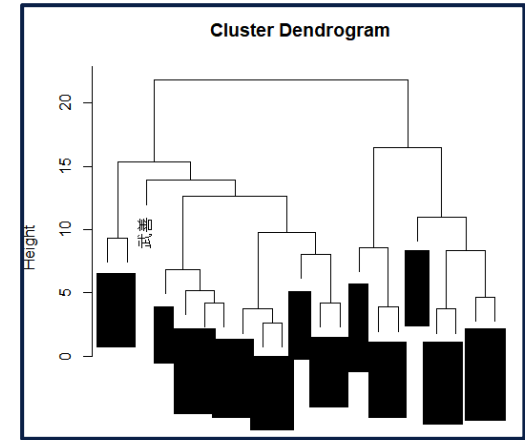
散布図行列

情報Ⅱ そのまま見る!

主成分分析で、次元を圧縮しよう



クラスタリングで、グループ化しよう



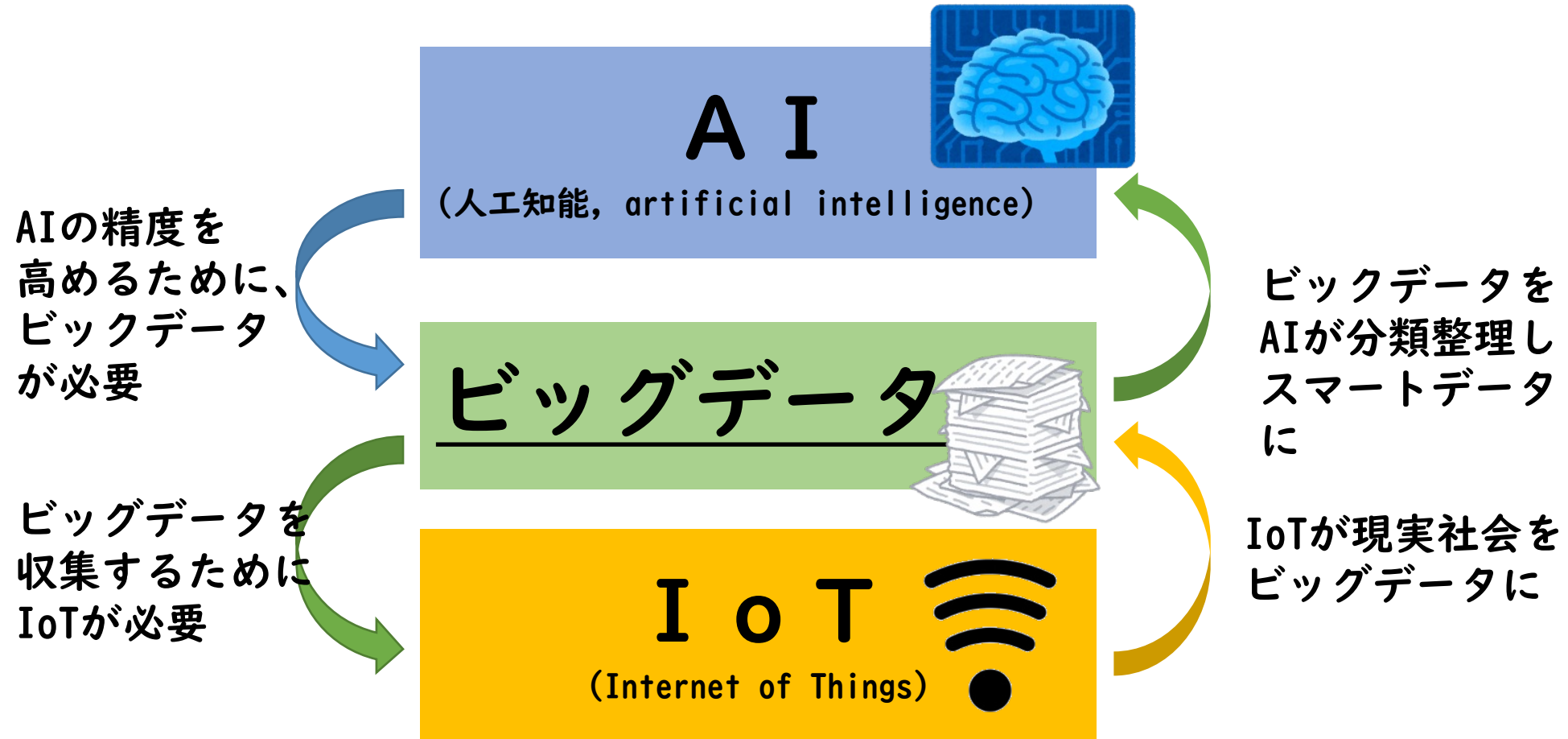
etc...

Agenda.

- 1 情報Ⅰ／情報Ⅱのデータサイエンス(学習指導要領より)
- 2 情報Ⅱにおける実践
- 3 情報Ⅰにおける実践
- 4 データサイエンスを扱う価値3つ(個人的に)**

①社会の在り方 今後の社会の仕組み(データ駆動型社会)

データは21世紀の**石油**である。



以下Webページの図を参考に作成：

<https://it.impress.co.jp/articles/-/14382>

②情報科の要 「試行錯誤」の場面が豊富なのもデータサイエンス

多くの生徒が、勉強のせいで睡眠不足に陥っている！今後、改善したい！
他に考えられる結論は？

結論
Conclusion

C

分析
Analysis

A

P

問題の決定
Problem

P

計画
Plan

D

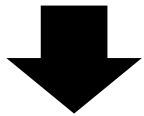
データ収集
Data

睡眠時間と学習時間に関係性があるか？
類似するテーマには何がある？

クラスに、アンケート調査をやってみよう

アンケートを実施
アンケートはどのように取る？

情報Ⅱで視野が大きく広がる！



学習時間が多い人は、睡眠時間が短い！
分析手法を変えてみよう！

③統計(データサイエンス)は人生を豊かにする

「雨乞いをするとう雨が降る」……？

	降雨あり	降雨なし
雨乞いあり	A	B
雨乞いなし	C	D

「雨乞いをするとう雨が降る」
ためには、
「雨乞いをしないと雨が降ら
ない」
ことも必要

人間

自分の信じたい仮説を支持する
情報を積極的に集めてしまう
(**確認バイアス**)

→素早い情報処理

コンピュータ・統計

感情に左右されず、数学的に物
事を処理していく。

→綿密な情報処理

統計はパートナー。世の中の見え方を変えてくれる

まとめ

- 情報Ⅱは、案外手を出しやすい。
- 「おっ、今度はこのデータを分析してみたいぞ！」と思えるのは情報Ⅱの方が多いかも。
- ある高校生の言葉
「高校生活で一番感動したことは、主成分分析です」

終

発表資料は全てWebで公開しています

<https://www.hinode.ed.jp/share/takeyoshi/>