

# テキストマイニングによるアンケート分析

## —Reborn-Art Festival 2021-2022を事例として—

浅沼大樹

### 1. はじめに

#### 1.1 Reborn-Art-Festival について

石巻専修大学の所在地である宮城県石巻市では、2011年の東日本大震災による被災からの復興過程において、さまざまな取り組みが行われてきた。中でも、石巻市の中心部から牡鹿半島および女川町にかけての広範囲のエリアを対象として開催されるReborn-Art-Festival(以下RAF)は一つの象徴的なイベントと言えるだろう。RAFは2015年に石巻市とap bank<sup>1</sup>が发起人となりRAF実行委員会が発足し、2017年に初回の本祭が開催され、その後2年に一度のペースで実施されている。

RAF2021-2022は直近に開催された3度目の本祭であるが、COVID-19の流行期に重なっており会期中に緊急事態宣言が発令されるなど、感染症対策に対して国民が最も敏感になった時期での開催となった。そのため、本来のタイミングとしては2021の本祭開催となるところが、規模を縮小して前期(2021年開催、以下RAF前期)と後期(2022年開催、以下RAF後期)に分割した形で開催されることになった。なお、RAF前期には約10万人の来場があり宮城県全体への経済効果は約6億4,140万円、RAF後期は約12万人の来場があり経済効果は約6億3,580万円と報告されている<sup>2</sup>。

#### 1.2 本稿の目的

さて、このように人の動きや経済の動きについて石巻市に大きな影響を与えているRAFであるが、その開催時に実施される来場者へのアンケートは、事務局である一般社団法人Ishinomaki 2.0が所管している。オンライン・オフライン両方を含めたアンケート結果がIshinomaki 2.0によって取りまとめられているものの、残念ながらその分析については十分に行われているとは言い難い。

前期・後期合わせて約2,000件のアンケート結果が得られているが、集計は各項目についての単純集計にとどまっており、自由記述については特に掘り下げた集計・分析もされていない状況である。一方、単純集計を超えるアンケート分析には一定の専門知識やツールの取り扱いが必要であり、他に数々の業務を抱えたIshinomaki 2.0がそれを十分に行うことが難しい現状もある。そこで、本稿ではExcelとpythonを用いたテキストマイニングの手法を用いて、RAF前期および後期のアンケート分析に取り組むことにする。なお、アンケート結果の利用につい

<sup>1</sup> 音楽家の小林武史氏および坂本龍一氏と Mr.Children の櫻井和寿氏によって2003年に設立された一般社団法人である。設立当初から現在まで、環境保全や復興支援などの活動を継続的に行っている。<https://www.apbank.jp/about/>

<sup>2</sup> Reborn-Art-Festival 経済効果推計報告書より。

てはIshinomaki 2.0の松村代表から許可をいただいた上で分析を行なっている。

### 1.3 本稿の構成

本稿の構成は以下のとおりである。続く第2節ではアンケート分析に関する先行研究とその手法などについて考察する。データを使った実際の分析と考察の内容を第3節で記述している。また、今後の課題についても明らかになるだろう。最後に第4節は結論である。

## 2. 先行研究

コンピュータで文章を取り扱うことができるようになり、アンケートの自由記述の内容や作家が書いた文章の解析を行うことが容易になっている。また、人工知能を駆使したビッグデータの分析から意味のある関係性を引き出す「データマイニング」と呼ばれる手法の隆盛も手伝って、コンピュータでテキストデータの解釈を行う「テキストマイニング」が学術研究のみならず、企業のマーケティング手法の一つとしても普及しつつある。テキストマイニングで定評があるのはKH Coderと呼ばれるソフトであり、Excelと組み合わせることで文章をコンピュータが解釈可能な形に変換することができる。ソフト開発者の樋口耕一氏によれば、「従来の社会調査においては、数値化されていない文章のようなデータ、すなわち質的データの利用には種々の困難があった。・・・計量テキスト分析の特徴は、『テキストマイニング』とよばれる比較的新しい技術を活用しつつ、伝統的な内容分析(content analysis)の考え方を実践に活かすこと」であり<sup>3</sup>、現在ではそのような手法はKH CoderのみならずPythonのような汎用プログラミング言語でも充実してきている。本稿では、主に石田(2022)のテキストを参考にしつつ、アンケート分析を実施した。

そのような背景もあり、アンケート分析をテーマとした研究論文は数多く出版されている。例えば、西村・清水(2021)では講習会の受講者が講習をどの程度理解したかという理解度をアンケートの自由記述の分析によって明らかにしている。また、越中他(2015)や永田(2022)も同様に講座の内容についてアンケート分析を行い、受講者の理解度把握を試みている。特に、永田(2022)は本稿と同様Pythonを使ったテキストマイニングを実施しており、本稿の分析手法を確認する上でも非常に参考になった。

さらに、RAFには地域活性化に対する役割も期待されている点は指摘しておかなければならない。地方におけるアートプロジェクトが地域活性化に資する役割を果たしている事例は例えば室井(2017)や三浦(2017)で分析されている。また、金光(2018)はアートフェスティバルのアンケート分析を通じて、イベントが地域住民にどのような認識の変化を与えるのかという視点からの分析を行なっている。その中では開催地域によって地元住民の考え方が異なる可能性が指摘されており、RAFの地元住民の中での位置付けをどう考えるか、など今後のアートのあり方を考える上でも示唆に富むものであると言える。アートプロジェクトとして国内で有名なものはなんといっても香川県の瀬戸内国際芸術祭であろう。火付け役のこのイベントを

3 樋口(2014) p.1より引用。KH Coderを使った分析の入門書では、樋口(2014)のほか、例えば末吉(2019)などを参照のこと。

契機として、全国各地で様々なアートフェスティバルが開催されている。研究も徐々に蓄積されつつあり、瀬戸内国際芸術祭を取り上げる学術論文も多くある。上述の室井(2017)と金光(2018)も分析対象としているのは瀬戸内国際芸術祭である。

RAFもこうした「アート×地域」の流れの中に位置付けられるものであるが、残念ながらまだきちんとした分析の対象とはなっていない。そこで、まずは試論的に本稿においてアンケート分析を行い、今後の研究へと繋げていくことを試みる。

### 3. アンケート結果の分析

#### 3.1 属性に基づくクロス集計

本節では参加者の性別や居住地といった属性と満足度の関連について調べるために、クロス集計の結果を示す。まず、性別と満足度に関するクロス集計の結果をまとめたのが下の表1および表2である。表1がRAF前期、表2が後期のものとなっている。総回答数は前期が378、後期が1474となっており、前期・後期どちらも女性の方が男性よりも1.5倍ほど回答数が多い。回答結果は「大変満足・満足・やや満足」といった満足感を示す回答が大半を占める。一方で、不満感(大変不満・不満・やや不満)を表明する回答も少数ではあるが存在する。

さらに、この回答内容に性別による差があるのかどうか確認するため、クロス集計の結果をカイ2乗検定にかけた。その結果、前期についてはp値が $0.041 < 0.05$ であり、「回答内容に性別による差は無い」とする帰無仮説は棄却された(有意水準5%)。したがって、前期においては満足度について統計的に有意な性別による差があったと結論づけられる。一方、後期のクロス集計ではp値は $0.95 > 0.05$ となり、帰無仮説は棄却されなかった。したがって、後期においては満足度に性別による差はなかったということになる。

表 1 性別と満足度に関するクロス集計【前期】

性別	大変不満	不満	やや不満	どちらとも言えない	やや満足	満足	大変満足	合計
女性	3	8	1	7	35	97	77	228
男性	1	7	7	9	29	50	47	150
合計	4	15	8	16	64	147	124	378

表 2 性別と満足度に関するクロス集計【後期】

性別	大変不満	不満	やや不満	どちらとも言えない	やや満足	満足	大変満足	合計
女性	2	2	7	28	120	429	296	884
男性	1	1	5	16	69	295	203	590
合計	3	3	12	44	189	724	499	1474

次に、居住地に関するクロス集計の結果を示す。RAFには全国各地から多くの訪問者があるが、表3・表4を見るとわかるように、アンケート回答者の中で特に多いのは、前期・後期ともに石巻市、石巻市以外の宮城県内、東京都である<sup>4</sup>。なお、性別のクロス集計と合計値が

4 アンケートの取り方が前期と後期で異なっており、前期は宮城県(石巻・女川以外)となっている一方、後期では宮城県(石巻市以外)という項目立てがされているため、厳密には両者は比較可能ではないが、ここでは宮城県居住者は同じカテゴリとして扱うことにする。

一致しないのは、性別・居住地のどちらにも無回答者が一定数存在するからである。

アンケートのクロス集計には、属性と満足度のどちらにも回答があるものしか使うことができないので、性別・居住地・満足度のどれかの項目に無回答がある場合は、その回答を削除して集計している。したがって、(性別・満足度)のクロス集計の合計値と(居住地・満足度)のクロス集計の合計値では、自ずと値に差が出ることになる。

表 3 居住地と満足度に関するクロス集計【前期】

居住地	大変不満	不満	やや不満	どちらとも言えない	やや満足	満足	大変満足	合計
北海道	0	0	0	0	0	1	1	2
青森	0	0	0	0	0	0	1	1
岩手	0	0	0	0	1	1	1	3
山形	0	0	0	1	1	2	2	6
福島	0	0	2	0	2	3	3	10
栃木	0	0	0	1	0	0	1	2
群馬	0	1	0	0	0	3	2	6
埼玉	0	0	0	0	4	3	2	9
千葉	0	1	1	0	0	5	4	11
東京	0	2	0	2	11	23	22	60
神奈川	0	0	0	0	3	6	2	11
富山	0	0	0	0	1	0	0	1
石川	0	0	0	0	0	0	1	1
岐阜	0	0	0	0	0	1	0	1
静岡	0	0	0	1	0	0	0	1
愛知	0	0	0	0	0	4	1	5
京都	0	0	0	0	0	1	1	2
大阪	0	0	0	0	0	5	2	7
兵庫	0	0	0	0	1	0	1	2
山口	0	0	0	0	0	0	1	1
高知	0	0	0	0	0	0	1	1
福岡	0	0	0	0	1	0	0	1
大分	0	0	0	0	0	1	0	1
宮城県(石巻・女川以外)	4	11	5	9	32	71	66	198
女川町	1	0	0	0	0	0	0	1
石巻市	0	0	0	2	8	19	10	39
合計	5	15	8	16	65	149	125	383

さて、居住地と満足度の関係であるが、こちらも性別と満足度の関係と同様に、満足感(大変満足・満足・やや満足)という感想を抱く回答が圧倒的に多いことがわかる。ただ、宮城県(石巻・女川以外)からの参加者については不満感(大変不満・不満・やや不満)の回答も少数ながら存在する。

しかし、居住地別に(石巻・東京)、(宮城・東京)、(石巻・宮城)とそれぞれ属性に分けてカイ2乗検定にかけた結果では、どの組み合わせにおいても居住地による回答に統計的に有意な違いは存在しないことが示された。したがって、ここに現れている不満感を示す回答は、単に開催地域以外の宮城県内からの参加者でアンケートの回答数が多かったという事情によるものであると判断できる。

表 4 居住地と満足度に関するクロス集計【後期】

	大変不満	不満	やや不満	どちらとも言えない	やや満足	満足	大変満足	合計
北海道	0	0	0	0	0	3	0	3
青森	0	0	0	1	0	1	3	5
岩手	0	0	0	0	3	14	11	28
秋田	0	0	0	0	0	5	1	6
山形	0	0	0	1	6	14	7	28
福島	0	0	0	0	3	5	10	18
茨城	0	0	0	0	4	7	2	13
栃木	0	0	0	0	1	3	0	4
群馬	0	0	0	0	0	3	3	6
埼玉	0	0	0	0	7	12	8	27
千葉	0	0	0	0	5	13	11	29
東京	0	0	2	7	20	72	56	157
神奈川	1	0	0	2	4	21	8	36
新潟	0	0	0	0	0	0	1	1
富山	0	0	0	0	0	0	2	2
石川	0	0	0	0	0	2	0	2
山梨	0	0	0	0	0	1	0	1
長野	0	0	0	0	0	0	1	1
岐阜	0	0	0	0	0	0	2	2
静岡	0	0	1	0	0	4	2	7
愛知	0	0	0	0	0	2	4	6
三重	0	0	0	0	0	0	1	1
京都	0	0	0	0	1	0	0	1
大阪	0	0	0	0	0	3	6	9
兵庫	0	0	0	0	0	3	0	3
奈良	0	0	0	0	0	1	0	1
岡山	0	0	0	0	0	0	2	2
香川	0	0	0	0	0	0	2	2
高知	0	0	0	0	0	0	1	1
福岡	0	0	0	0	0	0	1	1
宮城県(石巻以外)	2	3	7	28	103	434	278	855
国外	0	0	0	0	1	1	1	3
石巻市	1	0	2	9	36	124	83	255
合計	4	3	12	48	194	748	507	1516

### 3.2 自由記述のワードクラウドおよびコレスポネンダンス分析

次に、アンケート項目の自由記述内容についての分析を試みる。ここでは回答者の属性を、例えば[女性・20代]などのように年代と性別の組み合わせとして取り、その属性ごとに自由記述に書かれていることを自然言語処理の手法を用いて抽出し可視化する。

自由記述の回答内容は当然ながら日本語の文として作成されているので、Excelのような表計算ソフトではそのままの形では取り扱うことができない。しかし、自由記述の内容には非常に豊かな情報が含まれていることが多く、その分析はアンケートを実施することにおいて重要な意味を持つ。このような言語そのものを分析の対象として、作成された文章からその特徴を読み解く分析手法のことを一般に「テキストマイニング」といい、テキストマイニングのさまざまな手法が考案されてきた<sup>5</sup>。

自然言語処理を行う方法はいくつもあるが、本稿の分析には汎用プログラミング言語であるPython(パイソン)を使用した。Pythonは現在もっとも広範に利用されているプログラミング言語の一つである。Python自体も無償で利用できるが、さまざまなライブラリ(分析目的に合わせて使用することができるプログラムの枠組みのようなもの)も無償で利用することができ、自然言語処理に関するライブラリも充実しているため、こうした分析に適していると言える。

#### 【ワードクラウド】

テキストマイニングの一つの手法としてよく用いられるのが、ワードクラウドである。これ

5 ここには、コンピュータ科学における「自然言語処理」という研究分野が大きな貢献をしている。そして、この自然言語処理の手法を積極的に用いて作り上げられてきたのが、Chat-GPTをはじめとする生成系 AI と呼ばれる人工知能である。



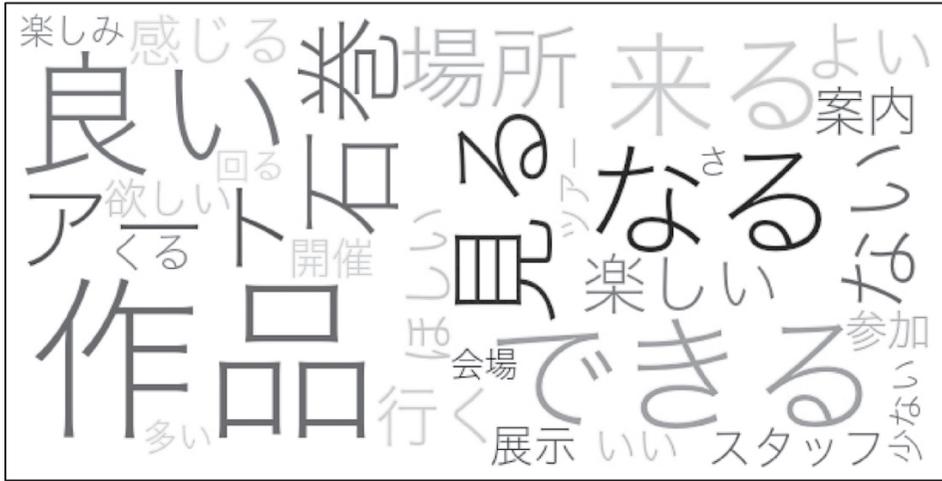


図2 RAF後期アンケート「自由記述」の内容から抽出したワードクラウド

ワードクラウドは自由記述欄に記述された内容を概観し直感的に理解する助けとなる手法である。しかし、同時にその描画内容から記述内容の特徴を把握するには情報が不足している。そこで、回答者の属性と回答内容の特徴を関連付ける手法として、「コレスポネンス分析」という手法がよく用いられる。下に示した図3～図6が「性別および年齢」と「居住地」を属性として自由記述の記述内容と関連付けたコレスポネンス分析の結果である。

【コレスポネンス分析】

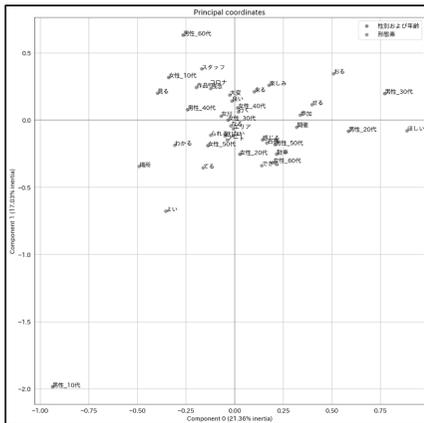


図3 RAF前期アンケート結果のコレスポネンス分析  
性別および年齢

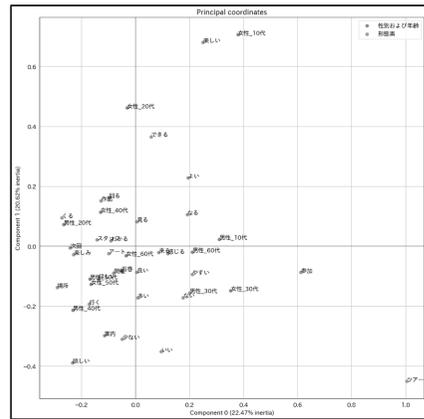


図4 RAF後期アンケート結果のコレスポネンス分析  
性別および年齢

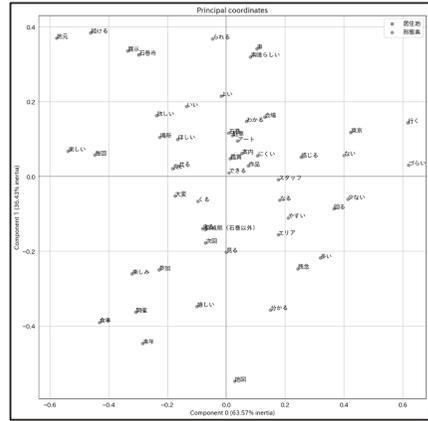
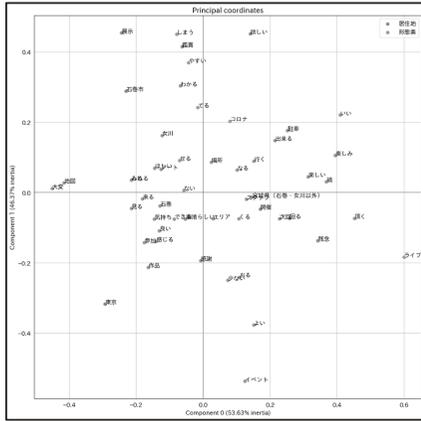


図 6 RAF 前期アンケート結果のコレスポネンス分析 居住地  
図 5 RAF 後期アンケート結果のコレスポネンス分析 居住地

コレスポネンス分析とは、いわば「クロス集計表の可視化」である。前掲表1、表2、表3の性別・居住地別のクロス集計表は各属性に含まれる人たちが回答した満足度の回答数を表にまとめたものであるのに対して、コレスポネンス分析の元になっているのは以下のような属性と「形態素」に関連するクロス集計表である。

表 5 [年齢-性別]と形態素のクロス集計表

	おる	せる	てる	できる	ない	なる	ほしい	よい	られる	わかる
女性_10代	0	0	0	0	0	1	0	0	1	0
女性_20代	2	3	1	9	5	6	3	6	5	1
女性_30代	4	4	4	14	9	14	4	2	9	7
女性_40代	8	10	6	15	12	30	6	4	8	6
女性_50代	5	4	9	19	11	17	2	11	5	11
女性_60代	1	1	3	3	7	2	1	1	1	1
男性_10代	0	0	0	1	0	0	0	1	0	0
男性_20代	0	1	0	1	1	1	5	1	1	0
男性_30代	8	5	0	7	2	3	5	0	0	0
男性_40代	2	3	5	6	6	6	2	7	3	4
男性_50代	2	0	3	8	2	4	2	1	1	1
男性_60代	4	1	0	1	4	2	0	0	1	1
	アート	エリア	コロナ	スタッフ	作品	参加	場所	大変	女川	感じる
女性_10代	1	0	1	0	2	0	0	0	0	0
女性_20代	3	3	1	3	10	2	3	3	2	4
女性_30代	11	7	6	9	18	6	7	5	12	8
女性_40代	15	14	15	9	34	9	7	11	8	4
女性_50代	19	7	9	6	31	5	15	4	4	6
女性_60代	0	0	0	0	1	1	0	0	0	1
男性_10代	0	0	0	0	0	0	1	0	0	0
男性_20代	3	0	1	0	3	0	0	2	1	3
男性_30代	2	1	3	2	6	5	0	2	0	1
男性_40代	6	2	10	4	23	3	6	8	2	2
男性_50代	4	4	2	2	6	0	0	0	3	2
男性_60代	1	1	4	7	19	0	2	2	2	2
	来る	楽しみ	欲しい	残念	石巻	良い	行く	見る	開催	駐車
女性_10代	0	0	0	0	0	0	1	1	1	0
女性_20代	3	0	2	1	6	2	4	4	3	3
女性_30代	5	6	4	8	10	14	11	16	10	11
女性_40代	9	7	5	5	14	12	21	14	14	3
女性_50代	5	5	6	7	17	14	7	10	18	3
女性_60代	1	1	1	2	2	1	3	1	2	1
男性_10代	0	0	0	0	0	0	0	0	0	0
男性_20代	4	0	1	0	3	1	2	0	2	0
男性_30代	2	4	1	2	5	5	1	0	11	2
男性_40代	7	3	5	3	6	8	4	15	8	2
男性_50代	5	0	2	1	6	4	5	1	5	5
男性_60代	6	2	1	5	2	6	4	6	1	0

性別と年齢を属性として、自由記述の内容からこうしたクロス表を作成する。列名の項目として入っているのが、記述された内容から抽出された「形態素」である。一般に、自由記述から抽出された形態素の数は膨大になることが多い。したがって、前述のように抽出する形態素数は30としている。そして、今回は性別と年齢をもとに属性を定義しているので[女性10代]~[男性60代]まで属性の数は12となった。結果としてクロス集計表は12行×30列の行列となる。

コレスポネンシ分析は、この形態素の出現頻度を表す行列を可視化する分析手法である。具体的な仕組みについて詳述することは本稿の射程範囲ではないので避けるが、イメージとしては3次元の物体に光を当てて2次元の影を作り、その影を見て元の物体の情報を復元するようなものである。3次元以上は物理的に表現不可能だが、同じように2次元に次元を落とすことができれば平面のグラフに変換することができる。ただ、光の当て方により物体の影は形を変えてしまい、影の形が悪いと元の物体の情報を再現率が低いこともある。コレスポネンシ分析では、なるべく元の物体の形を再現した影を作るようにクロス表を変形してグラフ化している。

さて、実際にグラフを見てみよう。コレスポネンシ分析のグラフは、図3のように散布図となる。それぞれの点が属性と形態素に対応しており、横軸と縦軸はその点が持つ「類似度」を表すと解釈できる<sup>7</sup>。つまり、図上の点は互いにその距離が近いほど同じ特徴を持つと言える。例えば図3の右下の領域には[女性\_20代]・[女性\_60代]・[男性\_50代]が近接しており、これらの属性に含まれる回答が類似した特徴を持っていると言える。これらの属性の周りには[できる]・[感じる]・[石巻]・[駐車]などが配置されているので、これらの属性に含まれる回答者はこうした感想を抱いていたと解釈される。

### 【課題】

ただ、このようにコレスポネンシ分析を行なっても、今回のケースではなかなか回答に明確な特徴を把握することは難しい。実際、[性別\_年齢]という属性分類では、自由記述という質問の性質上、質問項目自体に制約がなく、回答者も自身の意見を自由に述べているので、集計した時に特徴を把握するための軸となるものが見つけにくいからであると考えられる。また、形態素を抽出する際のプログラム上の設定を改善する余地も多分にあると思われる。

こうした課題を解決するには、アンケート作成の際の質問項目の設定を現行よりも方向性を持たせるようにすることが考えられる。例えば、全般的な自由記述欄を設ける前に「RAFの運営に関する要望」を聞き取る設問項目や「会期中の移動に関して気になる点」を抽出するための設問項目など、分析目的を事前に策定した上でそれに合致する設問項目をアンケート内に入れていくようにすれば、回答から特徴量を適切に引き出すことが可能になるであろう。した

7 図3の横軸と縦軸に記載されている inertia (イナーシャ) はその軸で再現可能な情報量を表している。両軸合わせて約 39% となるが、この数値は図3に含まれる情報量は元々のクロス表が持っていた情報の約 39% を再現しているという意味になる。もともなるクロス表のサイズがかなり大きいので、それを二次元に圧縮するとどうしても情報の欠落は防げない。一般的にコレスポネンシ分析のイナーシャは 40%~50% 程度が目安とされており、今回の分析ではそこまでの精度は出せていない。

がって、アンケートの作成はその分析結果の出力イメージからの逆算で作成するという視点を持って運営に当たる必要があるだろう。

### 3.3 アート評価の年代・性別によるクロス集計

最後に、アート評価のクロス集計を行う。RAFには毎回さまざまなアーティストが作品を展示する。そのどれもが個性的で、意図を持って作られた作品であり、量的な評価には馴染まないという側面は確かにある。また、作品の配置はアーティストの感性との合致が必要であるから、運営側が作品の配置を任意に定めるといってもいかならない。

しかし、一方で一般参加者からの人気が高い作品というのも確かにあり、そのような作品をどのように配置するかということは、特にRAFのような広い会場にアートを点在させるスタイルのアートフェスティバルにおいては戦略的に重要な事項となる。人気作品に人が集中することを考えると、会場範囲の回遊ルートや設計などにこうした視点は不可欠であると思われる。

したがって、RAF前期・後期を通じて、参加者からの評価(印象的な作品を最大3つ選ぶ)を属性ごとに取りまとめて評価を試みた。結果は表6・表7として本稿の最後のページに掲示している。各年代・性別属性をアンケート結果から抽出し、印象的なアート作品に対する回答ごとに回答の頻度を計算したものである。最大3つ選ぶということで、各属性の回答者が選んだ数が多かった作品3つを色分けして示してある。最後の行は合計値であり、ここも上位3つを色分けして示してある。この行はいわば回答者全体での評価を表していると言えるだろう。

一見してわかるように、RAFの象徴ともいべき名和晃平氏のWhite Deerは前期・後期通して人気が高い。この作品は2017年に開催された初のRAF本祭で披露されたのち、常設展示のアートとして人気が高い作品である。また、RAFの発起人でもあり実行委員長を務める小林武史氏が関係する作品も会期を問わず印象度が高いことがわかる。

一方、前期と後期の顕著な違いとしては、前期では前2作品以外の回答者の評価は比較的分散しているのに対し、後期は前2作品とともに小谷元彦氏の作品を印象的であると答えた回答者が多い。小谷氏の作品『サーフ・エンジェル(仮設のモニュメント2)』はRAF後期のパンフレットの表紙を飾った作品でもあり、RAF後期の実施前から注目されていた作品と言える。

## 4. 結論

ここまで、RAF前期・後期に実施されたアンケート結果の分析を行ってきた。主に属性別のクロス集計をもとにした分析であり、手法としては標準的なものばかりである。

RAF前期・後期で実施されたアンケートの回答はそもそも満足度が非常に高く、その意味では回答の多様性はそれほど高くないとも言え、どの属性でクロス集計を実施しても得られる結果はそれほど異なるものにはならない可能性が高い。とはいえ、居住地別のアンケート回答では石巻地域以外の宮城県民からの意見がことの他多く、その中には不満を表明するものも他

の地域からの参加者と比べて多かった<sup>8</sup>。属性と自由記述の関連をもう少し詳細に分析すれば、彼らが何に対して不満を持っているのか把握することができたかもしれないが、今回は本稿の目的を標準的な手法を用いて単純集計よりは詳細なアンケート分析を試みるという試論にとどめている。より詳しい分析は、今後実施していく研究の中で進めていくこととしたい。

また、この分析を通じて明らかになったことは、アンケートの設計についてももう少し詳細な検討が必要であろうということである。今後の実施運営に当たって回答者の気持ちや嗜好などを調査する必要があるのならば、それらを回答者が表明できるような形で設問事項を検討する必要がある。ワードクラウドやコレスポンデンス分析にかけたとき、そこに示される出力結果の特徴を解釈できるようにするためには、回答の枠組みを定めない自由記述欄のみでは回答内容にばらつきが大きすぎて、意味のあるアウトプットを出すのが難しくなってしまう。したがって、上述したように、自由記述のほかに要望を聞く設問や不便な点を聞く設問など運営の改善につながる回答が得られるような設問事項をあらかじめ設定しておくべきであろう。

#### 参考文献

- 石田基宏 『Pythonで学ぶテキストマイニング入門』 C&R研究所2022年。  
金光 淳 アート・フェスティバルは地域をどのように表象し何を可視化するのか：鳥連想イメージネットワーク分析 理論と方法33巻1号 数理社会学会 pp:114-131 2018年  
越中康治・高田淑子・木下英俊・安藤昭伸・高橋 潔・田幡憲一・岡 正明・石澤公明 テキストマイニングによる授業評価アンケートの分析—共起ネットワークによる自由記述の可視化の試み—宮城教育大学情報処理センター研究紀要 2015年  
末吉美喜 『テキストマイニング入門 ExcelとKH Coderでわかるデータ分析』 オーム社 2019年  
永田義久 2022年度後期福岡大学市民カレッジ「ヨーロッパの街と歴史をたどる」第一講のアンケート分析 福岡大学人文学論叢 第55巻第4号 pp:1065-1118 2024年  
西村奏咲・清水忠 テキストマイニングを用いたアンケート解析 薬学教育第5巻 日本薬学教育学会 pp:1-5 2021年  
樋口耕一 『社会調査のための計量テキスト分析 内容分析の継承と発展を目指して』 ナカニシヤ出版 2014年  
室井研二 過疎地のアートプロジェクトと地域活性化((三浦他(2017)pp:40-56所収)  
三浦典子 企業の社会貢献と地域再生 アートがつなぐ官民の力(三浦他(2017)pp:2-27所収)  
三浦典子・横田尚俊・速水聖子 『地域再生の社会学』 学文社2017年

---

8 もっとも、これはカイ2乗検定の結果でも指摘したとおり、属性に起因するものではない。単に宮城県民の回答数が他の地域と比べて多いという事実を反映しているだけである。

表 6 性別\_年代による印象的なアートの評価【前期】

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A	B1	B2	B3	C1	C2	C3	C4	C5	C6	C	C	C	C	D1	D2	D3	D4	D	E1	F	F	
	Hou xOQ ue	大友 xOQ	大友 xOQ	片山 真澄	雨宮 龍介	MES	アユ ンキ	片山 真澄	西原 謙之	高橋 聖久	増田セバ スチヤン	オ ノ コ	加藤 真	練田 太郎	さえ ポ	W X	岩根 愛	SWO ON	夏井 樹	益氏 龍平	久住 有生	小林 万里 子	片山 真澄	布野 樹太 郎	名和 亮平	名和理江子+ 藤田介+佐 藤美幸+菊池 聡太郎	吉澤 剛							
女性 10代	0	0	1	1	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
女性 20代	0	12	4	2	11	3	0	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
女性 30代	6	5	4	2	10	3	2	4	1	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
女性 40代	8	10	11	4	6	3	0	5	2	0	3	0	14	14	1	3	4	34	6	7	3	0	2	6	3	0	2	2	2	2	2	2	2	
女性 50代	1	2	0	0	0	0	0	0	0	0	0	0	9	9	1	2	3	22	3	7	1	5	4	5	2	1	24	0	3	1	1	1		
女性 60代	0	1	0	0	0	0	0	0	0	0	0	0	1	2	0	0	1	3	1	0	0	1	0	0	0	0	4	1	0	0	0	0		
女性 70代	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
女性 80代	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
男性 10代	4	4	4	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男性 20代	4	4	4	3	2	1	1	4	2	1	1	0	0	0	0	0	0	8	2	3	2	0	1	0	0	0	0	0	0	0	0	0	0	0
男性 30代	4	0	4	3	2	1	1	4	2	1	1	2	1	2	1	1	1	14	5	4	6	0	1	4	6	2	1	10	4	2	2	2	2	2
男性 40代	10	7	7	6	15	6	2	3	1	7	1	0	5	7	1	1	1	13	3	3	0	1	0	1	3	2	1	9	2	4	2	2	2	
男性 50代	3	2	0	8	1	6	2	1	9	0	1	0	4	5	1	3	2	13	3	3	0	1	0	1	0	1	1	9	2	4	2	2	2	
男性 60代	2	0	3	0	2	0	1	2	1	1	0	0	1	0	0	0	1	4	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男性 70代	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
合計	49	51	53	23	59	24	30	37	13	19	8	0	46	47	7	14	18	149	32	42	16	5	11	25	24	11	7	116	19	17	8	0	0	

表 7 性別\_年代による印象的なアートの評価【後編】

	A1	A2	A3	A	A5	A6	A7	A8	A9	A10	A10	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	C1	C2	D1	D2	D3	D1	D2	D3	D1	D2	D3	E1	E2	E3	
	山内 翔太	箱田 新子	アール フランス パキス ム	川原 正 忠	川原 正 忠	坂本 おる おる	カス タス メオ ス	船牧真理 子、与根 寛治	船牧真 理子	深瀬真 二郎	深瀬真 二郎	西宮 謙介	西宮 謙介	坂部 崇	坂部 崇	さわ ひら き	目[mile]	坂部 寛治	坂部 寛治	保良園 川原正 忠	SIDE CORE	小宮元 修	保良園 社	藤本 社介	小林 武史	伊勢 谷友 介	伊勢 谷友 介	久住 有生	金子 潤平	金子 潤平	有馬 かお る	有馬 かお る	有馬 かお る	有馬 かお る		
女性 10代	0	4	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
女性 20代	6	12	7	9	4	6	2	2	5	2	2	20	1	5	17	13	2	5	11	7	5	38	29	14	32	35	49	9	1	1	0	0	1	1	1	
女性 30代	12	14	5	9	5	7	5	5	6	3	24	2	3	9	11	0	3	7	13	5	47	28	29	23	80	6	1	3	3	6	10	6	5	10	10	
女性 40代	7	14	7	7	13	5	5	2	0	2	24	8	7	16	15	1	0	3	14	17	9	68	35	13	46	42	91	5	1	3	4	13	15	15	15	
女性 50代	7	9	9	8	9	3	3	3	5	0	16	8	9	12	13	9	11	8	25	7	55	45	15	52	31	72	9	1	3	9	10	13	13	13	13	
女性 60代	2	4	3	4	8	1	0	0	2	1	8	0	1	4	2	3	5	6	4	18	13	8	13	8	29	2	0	2	0	2	1	1	4	4	4	4
女性 70代	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男性 10代	1	0	3	1	1	0	0	0	0	0	2	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男性 20代	6	8	2	1	10	1	1	1	3	3	4	1	1	3	8	1	3	8	10	2	40	19	9	21	23	2	0	0	0	0	0	0	0	0	0	0
男性 30代	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
男性 40代	9	9	6	4	11	4	6	6	1	0	19	7	2	5	10	5	10	5	4	25	19	11	29	18	56	1	1	2	1	2	1	8	8	8	8	
男性 50代	5	7	2	6	9	7	7	1	2	15	3	2	6	8	3	5	8	7	7	42	24	9	25	14	38	1	2	0	1	2	0	4	4	7	7	
男性 60代	2	7	1	0	7	3	0	0	2	1	2	1	1	1	1	1	1	1	1	6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男性 70代	1	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
合計	59	97	46	5	82	45	35	35	27	16	153	36	84	97	31	43	81	114	50	423	261	96	235	530	41	8	21	34	59	73	0	0	0	0	0	