

時系列データに対する説明的可視化

牧 修平[○]（東京工業大学）

Explanatory Visualization for Time Series Data

Shuhei MAKI

ABSTRACT

Various visualization methods are used to make data easier to understand, but due to scalability and data structure, it may be difficult to recognize the features of the data we want to convey from the visualization. Therefore, in addition to visualization, research has been done to provide the features of data by words. In this paper, using the Social Progress Index, a time-series data with a large number of items as an example, we propose a system that provides visualizations to support data analysis and template-based summary sentences of the data. In this paper, we also explain how to search for data in the form of time series data and how to construct templates to generate summary text of the data. Finally, we show the usability of the system through a case study.

Keyword: Time Series Data, Explanatory Visualization

1. はじめに

生のデータを見るだけでは分析したり理解したりするのが難しいデータに対して、その分析や理解のために可視化を行う研究はたくさん行われてきている(Bryan, Ma, and Woodring 2016; Gschwandtner and Erhart 2018; Görtler et al. 2019; Sondag et al. 2020)。しかし可視化によるデータの分析や理解に不慣れなユーザや、可視化では特徴を正しく伝えきれないデータを分析したいユーザに対しては、データの可視化だけでその目的を達成することができない可能性がある(Ynneman, Löwgren, and Tibell 2018)。そこで文章と可視化の二つの方法を相互的に使ってデータの説明をする、説明的可視化が研究されてる(Hosokawa, Arimoto, and Wakita 2020; Latif and Beck 2019)。説明的可視化は、文章によってもデータの特徴を提供するため、可視化に精通するユーザーに対しても有益なアプローチである。

時系列データは日常の気温や降水量、近年では新型コロナ感謝数の推移など様々な場面で扱われているデータである。時系列データは必ずしも同様の事柄に対して一つのデータだけではなく複数のデータ項目が存在することがある(世界各国の新型コロナ感謝数の推移など)。このようにデータ項目数の多い時系列データは可視化による理解が困難になる。そのような時に、説明的可視化を使うことで可視化だけではなく、文章によってデータの特徴を伝えることができる。そこで本研究では [Social Progress Imperative](#) が 10 年にわたって発表している Social

Progress Index のデータ(Social Progress Imperative n.d.)を例として、説明的可視化による時系列データの理解と探

索を行う手法を提案する。以下 Social Progress Index のデータのことを SPI データと記述することにする。

2 章では説明的可視化と時系列データの可視化について関連研究をまとめ、3 章でシステムの設計について述べる。4 章ではシステムのユースケースを通してシステムの有効性を説明し、5 章と 6 章で考察とまとめを行う。

2. 関連研究

この章では、今まで研究してきた説明的可視化や時系列データに対する可視化をみていく。

Latif ら(Latif and Beck 2019)は、二変量の地理統計データを理解するために interactive Map Reports(以下 iMR)システムを開発し、説明的可視化を行っている。地図上の行政区の色や行政区に関する統計量を表すグリフを使って二つの地理統計変数を可視化するとともに、データを解析し

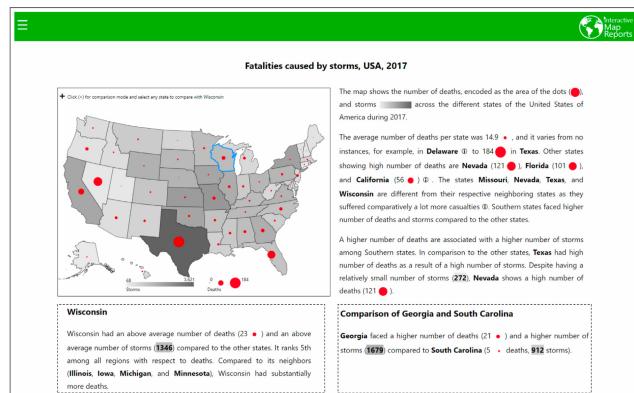


Figure 1: interactive Map Reports

て文章を自動生成しているのを、いくつかの例を上げながら説明している。図1はその例の一つであり、アメリカ合衆国の各州の年間の死者数とハリケーンの発生数に対してiMRのシステムを使用したものである。生成された文章には、「The average number of deaths per state was 14.9, ... (死者数の平均値が14.9万人だった。...)」などの統計的な特徴をはじめ、「Southern states faced higher number of deaths and storms compared to the other states.(南部の州は、他の州と比べて死者数やハリケーンの発生数が多い。)」など、特定の州のデータの特徴だけではなく、データの分布から見つかる地理的な偏りなどのデータの特徴も言葉で説明されている。iMRでは文章生成を行うプロセスを自身で定義し、ユーザーの要望に合わせてパラメータを変更することで柔軟な文章を自動で生成している。iMRは先で述べたようなデータの特徴を文章によって伝えることができるが、特徴が多い時にはシステムの画面からはみ出るほどの量の文章が生成されてしまい、ユーザーの負担になってしまう。

Latifらが開発したiMRは二変数のみの地理統計量データに対応していたが、Hosokawaら(Hosokawa, Arimoto, and Wakita 2020)は2019年のSPIデータを使い、データ項目数が多く、多変数の地理的統計量のデータに対して説明的可視化を行うSPIViewerを開発した。Hosokawaらはこのデータの特徴をユーザーに伝えるために、指標の階層性と国々の地理的階層を利用して文章を生成している。図2のでは、「機会」という指標とその子孫の指標について北アフリカの国々とサブサハラアフリカの国々を比較して文章を生成している。また、棒グラフを用いて一つの指標に対するデータの分布を表現したり、散布図を用いて二つの指標に対するデータの分布を比較したりすることができる。SPIViewerはSPIデータの階層性を活用してデータの特徴を文章によって伝えているが、時系列データに対してはサポートしていない。

時系列データの特徴を伝えるために可視化を使った研究は複数存在する(Andrienko et al. 2020; Bryan, Ma, and Woodring 2016; Van Wijk and Van Selow 1999)。そのなかでRoslingら(Rosling and Zhang 2011)は、データ項目数の多い二つの地理統計変数データを可視化するために、バブルチャートとアニメーションを用いたGapminderというシステムを開発した(図3)。時間に沿って動くバブルの位置や位置の変化によって、世界各国の二つの統計量の変化を提供している。しかしGapminderでは、どのデータの特徴に注目するかはユーザーに全て任せていた。そこでLuら(Lu et al. 2020)は、Gapminderのアニメーショ

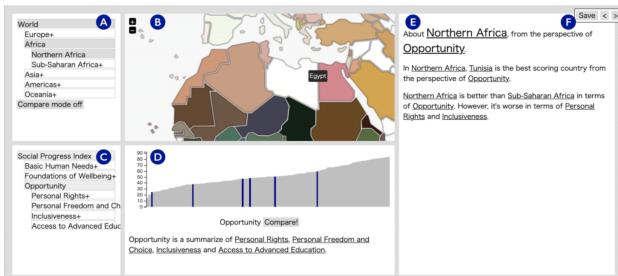


Figure 2: SPIViewer

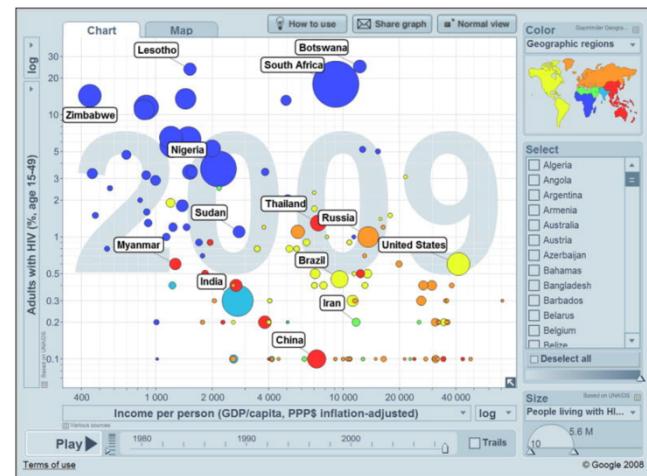


Figure 3: Gapminder

ンの中で注目すべきタイミングを抽出する研究を行った。Luらは時系列データの変化の傾向が変化した期間をデータ解析から特定して、時間を表すバーの中にその期間を表示している。RoslingやLuらの研究は、データを解析して誰かにデータの特徴を伝えるプレゼンテーションを目的にしているため、データの探索するためのサポートは行われていない。

本研究では、データ項目数の多い時系列データの特徴を可視化と文章にをユーザーに提示し、その後のデータの探索をサポートするシステムを設計する。文章生成において、Latifらを参考にしてテンプレートベースで文章を生成する。その時に文章の量が多くならないようHosokawaらのようにデータの構造に注目して文章を生成する。本研究では、SPIデータの国の地理的な階層性だけに注目して文章を生成しており、指標の階層性に注目した時系列データに対する文章生成は今後の研究課題とする。また、Luらが行った手法を参考に時系列データを解析して、時系列データの特徴を説明する。生成される文章の量を短くするために、解析する時系列データをユーザーの探索に合わせて選択する。

3. システムの設計

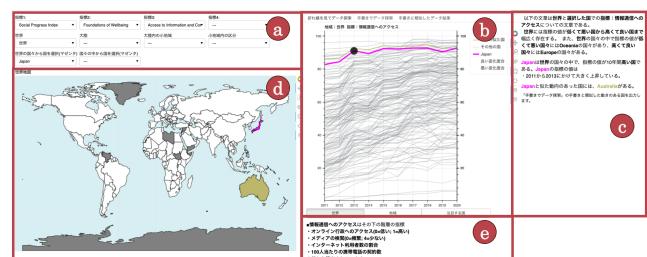


Figure 4: This figure shows the application of the system to the Social Progress Index dataset. In (a), the user can select the country, geographical area, and SPI index user wants to focus on. In (b) and (c), time series data visualization and automatically generated sentences are displayed according to the selection made in (a). In (d) is a world map to show the location of the country and geographical area of interest. (e) shows the definition of the SPI index to be selected in (a).

説明的可視化を通して、データ項目数の多い時系列データの探索をサポートするシステムを設計する。この章では課題からシステムの要件を定め、SPIデータを例にどのようにシステムを設計したのかを説明する。

3.1. 要件

データ項目数の多い時系列データに対する本研究におけるシステムの要件を定める。

R1 時系列データの変化の特徴: 同じ属性に対して時系列データが複数存在するとき、時系列データの値の変化は、他のデータの分布によって捉え方が変わる。例えば同じ1から2への変化でも、0から100の値を取る属性においては小さな変化、もしくはほとんど変化していないと捉えることができるが、0から10の値を取る属性においては大きな変化として捉えることができる。そこでデータを自動で解析して時系列データの変化に対する特徴を提供する必要がある。

R2 データの関係性から見た概要: データのなかには、データ項目間に階層性や包含など何かしらの関係が存在したり、データ項目に位置情報が関連づけられたりすることがある。そのようなときに、データの特徴を理解させるために、関係性やデータ項目に関連づけられた情報から見たデータの特徴も提供する必要がある。

R3 データ項目を絞ったデータの探索: データ全て解析してデータの特徴を全て可視化と文章で提示したとしても、ユーザーがその特徴を全て認識することができない。ユーザーが注目したいデータ項目だけに絞ったデータの探索をサポートする必要がある。

3.2. 設計

SPIデータを扱った例で、システムの設計を説明していく。

3.2.1. SPIデータ

SPIデータがどのようなデータか、どのような構造があるかを述べる。SPIデータは、Social Progress Indexという指標で世界各国を評価したもので、指標は70個ほど存在する。SPIの指標は、経済指標とは独立した実質的な生活の質の総合的な指標で世界各国を評価したものであり、社会の成功度合いを測るためにGDPなどの経済指標を補完するものとして設計されている。SPIデータでは、世界の各国をデータ項目としている。国はアジアやアフリカなどと州単位でまとめたり、東アジア、東南アジアなどといった州の中でも細分化された集合でまとめたりすることができる。つまりSPIデータにおいて、地理的範囲による階層構造の関係が存在すると考えることができる。しかし、SPIデータには地理的な階層の情報が含まれていないので、**国連**が提供している、地理的範囲の階層についてのデータ(UNSD 2003)を追加で扱う。

3.2.2. 可視化と文章生成

SPIデータをもとに、三つの要件に対してどのようにシステムを設計したか述べる。

データの選択: 探索する時に、ユーザーが注目したいデータ項目を絞れるようにする(R3)。SPIデータでは、注目

する国と指標に加え、先に述べた地理的範囲によるデータ項目の選択ができるようとする(図4a)。また、後に述べるインタラクションを導入することでこのデータ項目の選択を、より探索のニーズに沿って実現できるようにした。

傾向の変化する点: 時系列データの変化の特徴として、時系列データの傾向を調査する(R1)。そして時系列データの変化の傾向についてと、傾向の変化があるときには変化があった点をユーザーに提供する。まず時系列データ似たいとして線形回帰を行い、一次の直線を求める。その直線と元のデータに対して、全ての年でデータの差が閾値より小さい場合には単調な変化をしていると捉える。この閾値は指標に対する全てのデータ項目の値に依存しているため、この閾値を用いることでデータの変化が大きいか小さいかを判断することができる。変化が単調な場合は、線形回帰直線の傾きから、「10年間大きく上昇している。」や「10年間ほとんど横ばいである。」などといった変化の概要を捉える。一方どこかの年でデータの差が閾値以上であれば、傾向の変化がある可能性があるとみなす。次にJunhuaらと同様にWin法(Truong, Oudre, and Vayatis 2020)を活用して、データの傾向が変化する点を調査する。時系列データを (x_1, x_2, \dots, x_n) として、 (x_i, x_{i+1}) の線形回帰直線の傾き $m_i^{(1)}$ ($1 \leq i \leq n - 1$)と、 (x_i, x_{i+1}, x_{i+2}) の線形回帰直線の傾き $m_i^{(2)}$ ($1 \leq i \leq n - 2$)を求める。次に $d_i^{(j)} = |m_{i+1}^{(j)} - m_i^{(j)}|$ ($j = 1, 2$ $1 \leq i \leq n - 1 - j$)を求める。 $j = 1, 2$ 両方に対して、 $d_i^{(j)}$ が閾値以上であるとき、 x_{i+1} の点を時系列データの傾向が変化した点として判断する。この閾値も、指標に対する全てのデータ項目の値に依存している。この計算は、直近の左右のデータの傾きの差が大きい点をデータの傾向の変化があった点と見なしている。このとき、 $d_i^{(1)}$ は短い時間間隔でのデータの変化を、 $d_i^{(2)}$ は少し長い時間間隔でのデータの変化を捉えることができる。

類似しているデータ項目: 注目するデータ項目に対して、類似した変化と値をとっているデータ項目を探る(R1)。SPIデータでは、注目する国のデータに対してデータの変化が類似している国を求める。注目する国に対して、時系列データの相関係数が大きい国々の集合Cを集める。Cの中の国のデータ c_i と注目する国のデータでユークリッド距離 d_i を計算する。その距離 d_i がある閾値未満であれば、国 c_i を類似している国とする。この閾値も先ほどと同様に、指標に対する全てのデータ項目の値に依存している。二つの時系列データの上昇や減少などの大まかな変化が一致していることを相関係数によって絞り、注目する国と値が近く似た変動をする国を抽出するするために、時系列データのユークリッド距離を用いてる。

データの概要の可視化: 時系列データの詳細な特徴と、属性に対する全てのデータ項目のデータの分布をユーザーに伝えるために、折れ線を利用する(図4b)。データ項目数の多い場合では、多くのデータが重なって表示されるためにこの恩恵を受けることができない。データ項目数の多い折れ線からでも詳細なデータを提供できるように、特定のデータに対する折れ線をハイライトしたり、表示するデータを絞ったりする(R1, R3)。注目したい国と類似している国に対するデータをハイライトする。またデータの概要を提供するために、注目するデータ項目の時系列データの傾向が変化する点をドットを用いて表現しつつ、色を他の表現

と異なるマゼンタ色を使う。類似しているデータ項目に関してはダークカーキ色を使う。ドットの色は傾向の変化がその指標の中でいい変化であれば白を、悪い変化であれば黒をドットの色に使って表現する。複数傾向の変化があつた点が存在する場合、その変化の大きさをドットの大きさを使って表現する。

地理的範囲の表示: データ項目に関連づけられている位置情報をユーザーに提供する(R2)。SPI データでは、データ項目が世界各国に対応しているため、ユーザーが注目するデータ項目や地理的範囲を世界地図に表示する(図 4d)。注目する国と、類似している国については、それぞれ折れ線と同様の色で表現する。選択された地理的範囲に含まれていない国々と、データが存在しない国は灰色で表されている。これによって、ユーザーが注目する国と地理的範囲の国々とで地理的な関係を探索することができる(R3)。

文章生成: 可視化で伝えたいデータ特徴や、探索の手がかりとなるデータの特徴などを文章で伝える(図 4e)。文章は Latif らや Hosokawa らと同様に、テンプレートベースで生成する。文章は以下のデータ特徴について、データを解析して生成した。

- 選択した地理的範囲に含まれるデータ項目群が、世界から見て指標の値が低いか高いか(R2)。例: アメリカには世界の国々の中で社会進歩指標が高い国から高い国まで幅広く存在する。
- 選択した地理的範囲の子の階層の地理的範囲で、指標の値が高い、もしくは低いデータ項目が多いなどの偏りがあるか(R2)。例: アメリカの中では北アメリカに社会進歩指標が高い国が多い。
- 注目するデータ項目の指標に対する値が、選択した地理的範囲のデータ項目群の中で低いか高いか(R1,R2)。例: コスタリカはアメリカの国々の中では社会進歩指標が高い。
- 注目するデータ項目のデータはどのような動向があったか(R1)。例: コスタリカの社会進歩指標の値は、2011 年から 2020 年にかけてほとんど横ばいである。
- 選択した地理的範囲のデータ項目群の中で、注目するデータ項目と類似しているデータ項目について(R1)。例: コスタリカと似た動向のある国は、チリとウルグアイである。

3.2.3. インタラクション

ユーザーが可視化や文章からデータを探索できるように、インタラクションを導入する。

ズーム: 折れ線の表示において、地理的範囲のデータ項目群や注目したいデータ項目のデータの変動をより詳細に観察できるように、ボタンによって縦軸の範囲を調整できるようにしている。世界のデータの範囲を縦軸にすることで、地理的範囲に含まれるデータ項目群のデータが、世界と比べてどのような値をとっているか、その分布を確認す

ることができる。注目するデータ項目のデータだけに興味があるときは、そのデータの詳細が確認できる大きさにグラフの縦軸を調整する。選択した地理的範囲のデータ項目のデータの範囲を見ることは、データの分布が地理的範囲によって偏りが大きい時に役立つ。

フィルタリング: 折れ線の表示において、表示する折れ線をフィルタリングする機能を実装した。注目するデータ項目、類似しているデータ項目、その他のデータ項目のデータそれぞれで表示するか非表示にするか選ぶことができる。類似しているデータ項目が多いときにそれらを非表示にすることにより、注目するデータ項目と類似していない国を調べることができる。一方、類似しているデータ項目に注目したいときには、類似していないデータ項目を非表示にことができる。

クエリ: 時系列データを見るとき、特徴的な変化から探索したい可能性がある。そこで、ユーザーのフリーハンドや用意した折れ線の形を入力としてデータを探索できるようにした。

関連付け: 折れ線、世界地図、文章と三つの方法でデータの特徴を提供しているが、同じデータ項目に関する折れ線や地図、文字は同じ色で表現している。これにより、同じデータ項目に関するデータの特徴の理解を促進できる。

データ項目の選択: ユーザーが欲しいデータの特徴入手できるように、いくつかの視点でデータ項目を選択できる機能を実装した。折れ線の表示においては折れ線を選択することで、世界地図からは国を選択することで、注目するデータ項目を変更できるようにしている。これにより、データの変動の仕方や地理的な位置から、興味のあるデータ項目の特徴を得ることができる。

4. ケーススタディ

この章では提案するシステムが三つの要件、R1 時系列データの変化の特徴、R2 データの関係性から見た概要、R3 データ項目を絞ったデータの探索を満たされていることを示す。

図 5 は地理的範囲を「南北アメリカ」、指標を「個人の権利」、注目するデータ項目を「アメリカ合衆国」にした

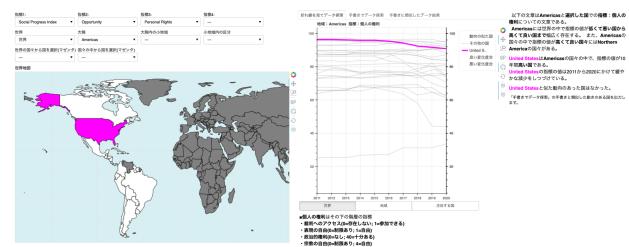


Figure 5: An example of using the system with Personal Rights as the index, the Americas as the geographical area, and the United States as the country of interest. It can be seen from the line chart that the United States has the highest value of Personal Rights in the Americas, and that it has been gradually decreasing for 10 years. It can also be seen that there are countries in the Americas with high and low values of the index, and that there are many countries in North America with high values of the index among the Americas.

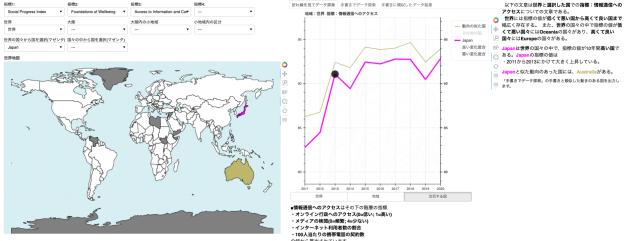


Figure 6: A comparison of data from Japan and Australia on access to information and communication shows that they are very similar.

例である。アメリカ合衆国は南北アメリカの中で最も個人の権利の値が高いこと、10年間徐々に減少していることが折れ線から読み取れる(R1)。文章の地理的階層に直した内容として、南北アメリカの国々は指標の値が高い国から低い国まで存在することや、南北アメリカの中で北アメリカには指標の値が高い国が多いことが書かれている(R2)。これらのことは、折れ線のデータの分布を確認したり、選択する地理的範囲を北アメリカに変更したりすることで確認することができる。次に注目する国に関する内容で、アメリカ合衆国が南北アメリカの中で指標の値が高いことや、10年間緩やかに減少していることがわかる(R3)。このことも折れ線のデータの変化や分布などから読み取ることができる。

他の国や指標についてもみていく。図4は地理的範囲を「世界」、指標を「情報通信へのアクセス」、注目するデータ項目を「日本」にしている。先ほどと異なってわかる特徴として、2013年に指標の値の変化の傾向が変わったことが折れ線からわかる。このとき、情報通信へのアクセスが伸びることは良いことなので、その上昇率が下がることはよくない変化として黒のドットに表されている。(R1)他には、類似している国に、オーストラリアがあることがわかる。類似していない国の表示を隠すと、日本とオーストラリアのデータは非常に似ていることが確認できる。(図6)次に文章の方をみていく(図4右)。先ほどと同様に世界の国々の中で、オセアニアには指標の値が低い国が多く、ヨーロッパには指標の値が高い国が多いことがわかる(R2)。国の特徴として、日本は2011年から2013年にかけてこの指標が大きく上昇したことを文章からも確認できる。

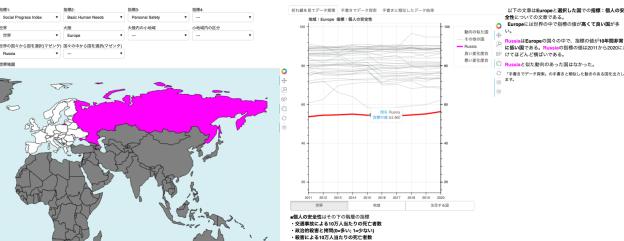


Figure 7a

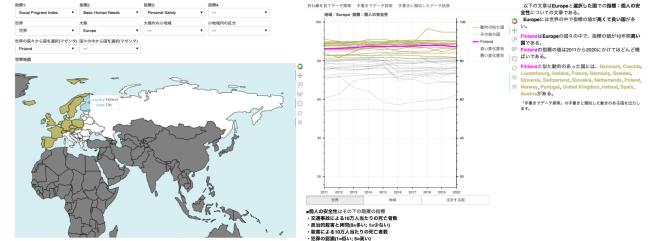


Figure 7b

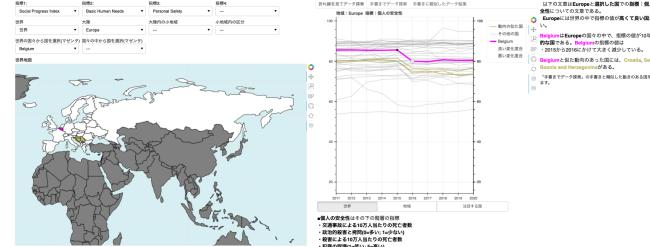


Figure 7c: An example where the indicator is personal safety and the geographic scope is Europe. The central figure in (a) shows that Russia has the lowest value for the indicator in Europe. In (b), the country of interest is Finland, and it can be seen that countries with similar data to Finland are clustered from west to north in Europe. Belgium, the only country in the west of Europe whose data was not similar to Finland's, showed a decrease in the value of the indicator from 2015 to 2016, as can be seen in the central figure in (c).

最後に別のアプローチでデータを探索していく。図7aは地理的範囲を「ヨーロッパ」、指標を「個人の安全」とした例である。まず折れ線をみるとデータが他のデータ項目より値が低いデータ項目があるのがわかる。ホバーすると、それがロシアであることがわかる。タップすると注目する国が「ロシア」に変わり、文章として詳細なデータの特徴を得ることができる。文章から、やはりロシアは指標の値がヨーロッパの中で非常に低いことがわかる(図7a右)。次にロシアと隣接している国は類似していない国ということから、それらの国データが気になる。ロシアの北西に位置する「フィンランド」を地図から選択してみる(図7b)。すると、ロシアとは異なりフィンランドは指標の値が高いことが折れ線の分布(図7b中央)からわかる。また地図(図7左)より、フィンランドとデータが類似している国がヨーロッパの西から北に集まっていることがわかる(R2)。しかし西側のベルギーという国は、フィンランドとデータが類似している国に含まれていない。そこでベルギーのデータを見てみると、ベルギー2015年までは周辺諸国と同じような値をとっていたが、2016年で大きく減少していることがわかる(R1)。

5. 考察

この章では3章で定めたシステム設計の三つの要件をどのように満たしているのか、システムの利点や欠点などを議論し、今後の研究課題について述べる。

5.1. 要件に対する充足

三つの要件に対してシステムを設計し、ケーススタディを通してシステムの有用性を示した。

R1 時系列データの変化の特徴: データを解析することで、データの傾向の変化する点を特定した。またその点を折れ線の中にドットして表現した。文章の中では、データの変化の傾向や、傾向が変化した点について述べた。

R2 データの関係性から見た概要: 個人の権利という指標で、北アメリカの国々が南北アメリカに対して値が高い国が多いことがわかるように、文章の中で地理的階層の特徴を提供した。

R3 データ項目を絞ったデータの探索: 注目するデータ項目、地理的範囲、指標に応じて可視化と文章を生成した。フィルタリングによって表示するデータ項目数を絞流ができるようにしている。また、注目するデータ項目の選択においては、データの形や地理的な位置から選択できるインターフェースを実装した。

5.2. システムの利点、問題点

次にこのシステムの利点と問題点を考察していく。

5.2.1. 利点

このシステムによって、データの分析における以下のようないくつかの利点がある。

- 元のデータからは見つけづらかった地理的階層に沿った特徴を、文章による概要から入手することができた。これは地理的範囲を変更することで実際にデータと示し合わせることができる。また、一度に複数の小さな地理的範囲の特徴を提供することができるのは有効だと考えている。
- 名前からデータ項目を選択するだけではなく、折れ線を用いたデータの形や分布、地図を用いた国お位置からデータ項目を選択することができた。データの形や注目している国と地理的に近い国のデータの関係などから、データを分析したくなることは自然であり、そのニーズに対応している。
- データの類似した国を折れ線や地図を通して提供することで、次の分析への手がかりになった。

5.2.2. 問題点

今回のシステムに利点がある一方で、以下のような問題点があった。

- 注目する国とその国に類似する国は地図上で色を変更したが、ガンビアのような国土面積が小さい国は色から場所を見つけることができない。現在は国全体を色付けしているが、小さな国に対しては、国を中心とした円を表示するなど別の表現方法が必要かもしれない。

- データの地理的分布の概要をみることができない。地図常にデータの値を表現する変数を実装していない。iMRのようにドットを用いて、大きさでデータの値を表すか国に使用している色で表すなどの変更が必要になる。しかし色を用いる際には、注目する国や類似した国をどのように表現するかが新たな問題となる。

- 選択している指標以外に対するデータの特徴がユーザにはほとんど与えられておらず、SPIデータ全体を通じた有益な探索と分析を行うことができない。Jinwookら(Seo and Shneiderman 2005)は多くの二次元散布図を相関やエントロピーなどさまざまな指標で評価し、ランクをつけて提供する研究を行っている。本研究のシステムでも、全ての指標に対して何かしらの基準で図を評価し、その結果をユーザーに提供する方法が有効かもしれない。

5.3. 今後の研究課題

このシステムには欠点がいくつか存在する。そこでこの節では今後の研究課題について述べる。

地図によるデータの可視化: 現在のシステムでは注目する国とその国と類似した国を示すために地図上の国の色を使用しているだけで、データの地理的分布がわからない。国の位置も大事だが、データの地理的分布にも興味があるため、この地図での可視化方法を再考する必要がある。一つの案は、iMRのように国之上にドットを配置し、その大きさでデータの値を表現するというものである。この場合、ドットの色を注目する国や類似している国などの意味を持つ色にすることができる。しかし、国が密集している範囲ではドットの大きさを正しく認識することができなくなることが予想される。もう一つの案は、今回実装した表示と、データの値に沿ったカラースケールの色で国を色付けした表示とを切り替えることができるようになることである。これによってユーザーのニーズに沿って効果的な表示を提供することができる。

テンプレートの変更: このシステムの文章には、時系列データに関しては注目する国データの変化特徴と類似する国特徴しか含んでいない。そこで今後は、指標に対する地理的範囲のデータが全体的にどのような変化があるかを文章にする。他に、文章を一つの指標に関してのみ生成したが、国のデータの多くが似ているような別の指標を提示したり、逆に全く似ていないような指標を提示したりするのも有効かもしれない。

分析の手がかりの提供: データが類似している国とその場所をユーザーに提供することで分析のサポートをしたが、有益な特徴を持っていそうなデータに対する指標を提示したり、複数の指標の関係を示したりといったことはしていない。文章にそのような探索の手がかりを提示することも有用だが、データの自由な分析のためには Jinwook らのように有益な特徴を持っていそうな表示を表示するビューが有用である可能性がある。今後はどのような基準で有用な

特徴をもったデータかを判断しつつ、ユーザーにそのことを提供することを考える。

6. まとめ

本論文では、可視化と文章によってデータ項目数の多い時系列データの理解や分析をサポートする説明的可視化を行うシステムを作ってきた。システムを作る上で、三つの要件 **R1 時系列データの変化の特徴**, **R2 データの関係性**から見た概要, **R3 データ項目を絞ったデータの探索**を定め、システムを作った。時系列の可視化は、データの詳細をユーザーが確認できるように折れ線を用いて表現した。このとき、データの変化の仕方からデータを探索できるように、手書きやあらかじめ用意したパターンからデータを探索できるインタラクションを導入した。データ項目は地理情報を含んでいるので、地図を用いることでデータアイテムの国位置を表示し、さらに地図上の国を選択することで、折れ線にその国詳しいデータを表示することができるようになした。一方、選択する指標、地理的範囲、国の視点からデータを解析し、地理的範囲の階層性に沿ったデータの特徴やデータの変化、類似した国について文章を生成した。その後のケーススタディではこのシステムの有用性をしました。文章を通してデータの概要を伝えることや、その後折れ線から詳細なデータを見て、事実を確認することができた。一方で、このシステムには問題点がいくつか見られた。可視化と文章によってデータの特徴を伝えることができたが、そのためにはユーザーによる探索が必要となる。しかし、本研究では探索の手がかりとなるデータの特徴を提示できていなかった。そこで今後は文章によって伝える内容の追加や、Jinwook らのように有益な特徴を持っていそうな指標を表示するビューを用意するなどの研究課題がある。

7. 謝辞

東京工業大学の脇田准教授には日頃の研究の助言に加え、論文の修正にあたり建設的な提案をしていただきました。ありがとうございます。

8. 参考文献

Andrienko, Natalia, Gennady Andrienko, Georg Fuchs, Aidan Slingsby, Cagatay Turkay, and Stefan Wrobel. 2020. "Visual Analytics for Data Scientists." In. Springer International Publishing.

Bryan, Chris, Kwan-Liu Ma, and Jonathan Woodring. 2016. "Temporal Summary Images: An Approach to Narrative Visualization via Interactive Annotation Generation and Placement." *IEEE Transactions on Visualization and Computer Graphics* 23 (1): 511–20.

Görtler, Jochen, Marc Spicker, Christoph Schulz, Daniel Weiskopf, and Oliver Deussen. 2019. "Stippling of 2d Scalar Fields." *IEEE Transactions on Visualization and Computer Graphics* 25 (6): 2193–2204.

Gschwandtner, T., and Oliver Erhart. 2018. "Know Your Enemy: Identifying Quality Problems of Time Series Data." *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 205–14.

Hosokawa, Natsuki, Kohei Arimoto, and Ken Wakita. 2020. "A Scalable "Exploramation" Technique for Hierarchically Indexed Table Data." In *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*, 1–5.

Latif, Shahid, and Fabian Beck. 2019. "Interactive Map Reports Summarizing Bivariate Geographic Data." *Visual Informatics* 3 (1): 27–37.

Lu, Junhua, Jie Wang, Hui Ye, Yuhui Gu, Zhiyu Ding, Mingliang Xu, and Wei Chen. 2020. "Illustrating Changes in Time-Series Data with Data Video." *IEEE Computer Graphics and Applications* 40 (2): 18–31.

Rosling, Hans, and Zhongxing Zhang. 2011. "Health Advocacy with Gapminder Animated Statistics." *Journal of Epidemiology and Global Health* 1 (1): 11.

Seo, Jinwook, and Ben Shneiderman. 2005. "A Rank-by-Feature Framework for Interactive Exploration of Multidimensional Data." *Information Visualization* 4 (2): 96–113.

Social Progress Imperative. n.d. "Social Progress Imperative." <https://www.socialprogress.org/>. Accessed: 2020-10-10. Accessed October 10, 2020. <https://www.socialprogress.org/>.

Sondag, Max, Wouter Meulemans, Christoph Schulz, Kevin Verbeek, Daniel Weiskopf, and Bettina Speckmann. 2020. "Uncertainty Treemaps." In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, 111–20. IEEE.

Truong, Charles, Laurent Oudre, and Nicolas Vayatis. 2020. "Selective Review of Offline Change Point Detection Methods." *Signal Processing* 167: 107299.

UNSD, UNdata. 2003. "United Nations Statistics Division."

Van Wijk, Jarke J, and Edward R Van Selow. 1999. "Cluster and Calendar Based Visualization of Time Series Data." In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*, 4–9. IEEE.

Ynnerman, A., J. Löwgren, and L. Tibell. 2018. "Exploramimation: A New Science Communication Paradigm." *IEEE Computer Graphics and Applications* 38 (3): 13–20.