

CA2031 「データ視覚化ロードショー」 ：大学図書館によるデータ視覚化講座

デヴィン R. ベイリー*
ジェフリー C. オリヴァー†
鎌田 均 (翻訳)‡

背景

視覚メディアを利用してデータからストーリーを効果的に語るための技術的スキルそしてデザインスキルへの需要が学術、産業分野にわたって増大している⁽¹⁾。データ量が増大し、また誤情報が一層蔓延している中で、こういったデータ視覚化スキルは重要となっている⁽²⁾。研究者は視覚化されたデータを利用するとともに作り出しており、そのため、視覚表現されたデータの読み取り方を理解し、人々を欺くことがないように効果的に伝えられるよう活用することが重要となる。データ視覚化を支援するサービスは、大学図書館の従来からの使命である情報リテラシー支援に合致し、米国を中心に近年比較的急速に拡大している。Cox らによれば、2016 年には 18%–22% の大学図書館がデータ視覚化の支援を提供しており⁽³⁾、5 年後の Zakaria の調査では 60% の大学図書館がワークショップやオンライン講座を通してデータ視覚化の支援を行っている⁽⁴⁾。データ視覚化のスキルが特定の学部学科に閉じられてサイロ化してしまうことがある一方で、大学図書館のような部門には、学内の幅広い対象にデータ視覚化支援のサービスを提供できる体制が備わっている⁽⁵⁾。

大学図書館によるデータ視覚化の支援には多くの事例が存在する。それらはデータ視覚化を支援する際に共通する課題に対処することを意識した内容となっている。こういった課題には、データセットの扱いに不慣れなこと、初心者にはデータ視覚化についての高度なコンセプトに触れる機会がないこと、学部教育初期の段階でデータを視覚化する機会が欠如していることなどがある。米・ニューヨーク大学健康科学図書館では、“Data Visualization Clinic” という、研究者が研究発表、文献で用いる視覚化について他者からフィードバックを受けることができる場を開設している⁽⁶⁾。この取り組みでは、研究者に自分のデータで作業する機会を与えることで、自分に馴染みのないデータを扱うことによる認知負荷を軽減させている。初級レベルでのデータ視覚化を支援することを目的とした別の事例では、データ視覚化のコンセプトを紹介する際に、一

般の Microsoft Excel[®]を用いている⁽⁷⁾。この事例での実践的アプローチでは、従来行われてきたようなデータ視覚化の理論の紹介を避け、初級の学習者がデータ視覚化スキルを継続的に伸ばしていくことを促している。また、大学図書館は、学部教育の初期段階に組み込んだ形で、データが重視される将来のキャリアに向けたデータ視覚化教育を支援することも可能である⁽⁸⁾。これらの事例において、データ視覚化スキルの向上を支援できる機会とそこに共通する課題が浮き彫りにされている。

筆者らが所属する米・アリゾナ大学では、データ視覚化のための支援がそれまでの受動的なものからより積極的なものへと近年転換された。これまでは、この支援は二つのサービス部門を通して主にその場その場で提供されてきた⁽⁹⁾。両サービス部門には、学内の研究者を支援するためにデータ視覚化及び分析応用の経験を持った人員が配置されており、筆者らの内、D. ベイリーは大学の情報技術サービス部門における“Research and Discovery Technologies”担当として、全学向けのサービス及び設備を介して研究支援を行っている。J. オリヴァーは大学図書館の部門である“Data Cooperative”に属し、研修、コンサルティングを通してデータに重きを置く研究を支援している。Data Cooperative のスタッフは大学院生、ポスドク、教員と連携して効果的なデータ分析、管理、視覚化のスキルを向上し、そのためのコンピュータ関係のリソースを整備する活動を行っている。

ワークショップ講座の要望が増加するなかで、筆者らは、データ視覚化支援に共通する課題に対処しつつ、短時間（1 時間）のワークショップ形式で教えることのできるデータ視覚化の具体的なスキルを特定することに取り組んだ。そこで筆者らが焦点を当てることにしたのは、データの視覚化が文字による伝達と比較して情報をいかに効果的に伝えることができるかについての理解と、効果的なデータ視覚化のためのソフトウェアの利用である。また、ワークショップの内容と提供方法を計画する上で、学習者に馴染みのないデータを使うこと（例えば土木工学者のためのワークショップに生物学のデータを使うこと）による余計な認知負荷と、初心者がワークショップ外で使うことが困難な過度に複雑なソフトウェアを使用することを避けるようにした。

サービスの実際

「データ視覚化ロードショー」と呼んでいる筆者らの取り組みは、初歩的なデータ視覚化の考え、ツール、大学内で提供されているリソースやイベントといった一連の内容を含んでいる。受講者が持っている知識に

*アリゾナ大学情報技術サービス

†アリゾナ大学図書館

‡京都ノートルダム女子大学国際言語文化学部

即した形で、受講者が日常会する場所で提供する、という双方の意味において「受講者がいるその場所で届ける」ために、実施には十分な時間をかけるようにしている。実施した講座は1週間以内に検討し、軽微な修正を加えている。例えば、講座を数回実施した後には、頻発し大きな変更を必要とするような課題点を確認するようにしている。これらを記録し、次学期の講座の予定を設定する前に対処している。以下の段落では、この講座の内容、計画、評価の概要について述べる。

講座は受講者と調整の上で日時を設定している。講座の機会を開拓する上では、授業を担当する教員とのメールによるルートが最もうまくいきやすい。もう一つのルートは知識、技術を共有するワークショップを計画している人物に接触することである。また、学部学科のプログラムコーディネーターやリエゾンライブラリアンと連携することも効果的であった。

講座は決められた枠組みに沿って実施されるが、その各部分は受講者に応じて変更される。全ての「データ視覚化ロードショー」は以下の内容を含んでいる。

- なぜデータを視覚化するのかについての議論
- ストーリーを語るためにどのようにデータの視覚化を利用するか
- 視覚化に利用できる一般的なリソース
- プログラム言語を用いたデータ視覚化の実践演習 (R⁽¹⁰⁾, Python⁽¹¹⁾)
- グラフィックツールを用いたデータ視覚化の実践演習 (Voyager⁽¹²⁾, RAWGraphs⁽¹³⁾)

データを視覚化する必要性を議論する際には、同一の情報をテキストデータのみによって表現した例とグラフィック表現の例を比較することが有効であった。筆者らがサービス開始初期に実施した土木工学の大学院セミナーでの講座では、合板組織の方向と対応する内部結合強さの表とグラフを提示した(図1)⁽¹⁴⁾。そして、内部結合から計測した合板の強度は繊維配向が直角の方が平行な時よりも大きいのか、受講者に質問した。テキストのみの表現(図1A)を見たときの受講者の一般的な回答は「たぶん」や少し躊躇した「はい」であったが、箱ひげ図(図1B)を見たときには、より確信を持って「はい」と回答していた。

A)

Orientation	Internal Bonding (MPa)
Perpendicular	1.68 ± 0.077
Parallel	1.26 ± 0.059

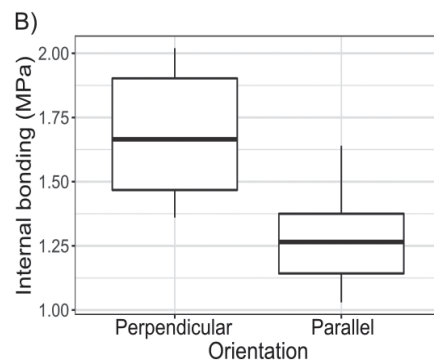


図1 (A) 表形式の繊維配向性質による合板の内部結合強さと変化、(B) グラフ形式の繊維配向性質による合板の内部結合強さと変化

この講座のサービスを開始してしばらくした頃からは、Tamara Munzner による「行動と対象の組み合わせ (Action, Target Pairs)」⁽¹⁵⁾を用いることにした。これは、視覚化された対象を見た際の一般的な行動のカテゴリーと、これらの行動の対象である元データにある要素を提示したものである(図2)。「組み合わせ」とは、動詞と名詞からなる、視覚化が意図するものを表す語の組み合わせを指す。例えば、視覚化の目的が期待分布から外れたデータポイントを識別することならば、行動と対象の組み合わせは「識別する、外れ値」となる。この例では、行動は「識別する」こととなり、対象は「外れ値」となる。

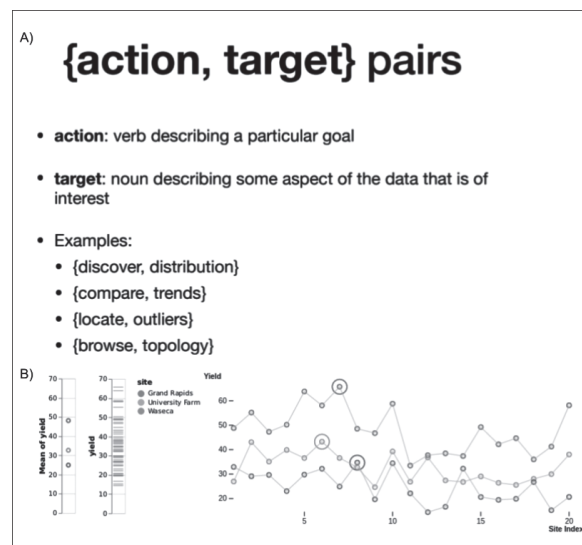


図2 (A) Tamara Munzner による“action, target pairs”の定義及び単語の例、(B) “action, target pairs”の理論を具体化した視覚化の例。左から右へ：比較する、傾向；発見する、分布；示す、最大値

こうしたデータ視覚化の理論についての講義の後に、双方向型の実践的活動に移る。これには3次元棒グラフ(図3)への批評が含まれ、例えば3次元目が何を付加しているか、なぜグループが二度描かれているのか、といった、グラフに潜在する問題点について受講者に質問している。

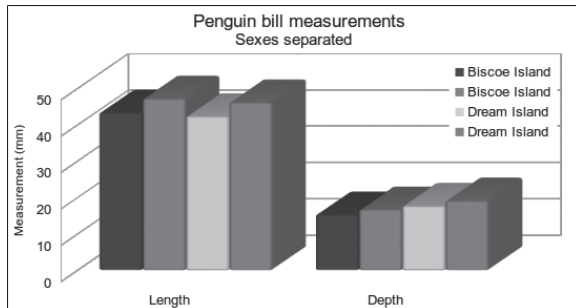


図3 データ視覚化を批評する双方向型セッションで使用した3次元棒グラフ

さらに後には、地理空間データの視覚化に詳しいアリゾナ大学図書館の同僚と協働するようになり、地図と対応するストーリーを選ぶ、“match the map to the story”という演習課題(図4)の提供をうけた。

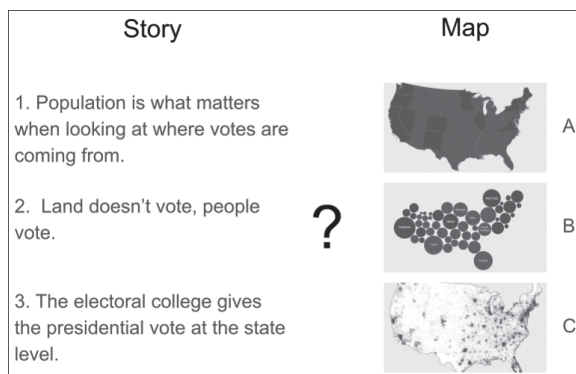


図4 双方向型セッションで使用したスライドの例。受講者に各々の地図をそれが語っていると思われるストーリーと対応させてもらった(B-1, C-2, A-3)。製作者 K. Carini による許可を得て複製。地図は Field⁽¹⁶⁾による。

これらの実践的活動では次の2点を強調している。まず、視覚化によってメッセージが効果的に伝わっているかを検討するために第三者から評価してもらうこと、そして、視覚化する方法の選択によって一つのデータセットが異なったストーリーを語り得る、ということである。

全てのワークショップ講座において、「データ視覚化を描く(Sketch a data visualization)」という受講者によるアクティビティを行っている。そこでは、5列から8列の小さなデータセットを見せて、受講者にこのデータを視覚化したものを手書きで作成してもら

う。グラフの種類についてわざと曖昧なままにしておくことで受講者が視覚化において創造性を発揮することを促している。講座のこの箇所は対象者に応じて変更した。例えば、環境科学の大学院生研究会での講座では土壌生化学のデータ(Kirsten Ball, unpublished data)を使用した。数学の大学院夏期研修では、SIR(susceptible, infected, recovered)感染モデル⁽¹⁷⁾、そして10万人あたりのCOVID-19感染者数とワクチン実施数を対応させたデータを使用した。筆者らが最初に実施した土木工学の学生対象の講座では、セメント濃度と圧縮強度との正の関係を表したテーブルを使用した⁽¹⁸⁾。このアクティビティでは、結局は伝えたいメッセージに適さなくなってしまうような視覚化に多大な時間を費やすのではなく、試作を迅速に作成することの重要性を強調した。

講座はこの時点で視覚化ツールを使った演習へと移る。ここでは、受講者に合ったデータセットと、BinderがホストするJupyter Notebooks⁽¹⁹⁾を使用し、受講者が簡単な演習課題を通してRもしくはPythonを使ってプロットを作成することを指導した。Jupyter NotebooksはRやPythonといった言語でコードを簡便に作成しテストできる環境を提供するものである。時間に余裕があり、受講者の関心が維持されていた場合には、別のツールObservable Notebooks⁽²⁰⁾を使った例も取り上げた。時系列データを視覚化する際のアニメーションの効果が分かる「棒グラフ競争」⁽²¹⁾は毎回受講者に好評であった。また、VoyagerやRawGraphsといったプログラミングを必要としないウェブでの視覚化ツールも使用した。これらのツールは、データを探求する上で視覚的に考えることを促し、とりわけ視覚化プロセスの初期に活用する際に有効である。

最後に利用可能なリソースを紹介することで「データ視覚化ロードショー」は締めくくられる。ここで筆者らが属する部門が提供するサービスを列挙し、連絡先を再度伝えている。アリゾナ大学図書館が毎年開催するデータ視覚化コンテストや、視覚化の主要な国際大会であるIEEE Vis⁽²²⁾のオンライン視聴といったイベント情報を提供している。また、受講者が取得した知識を実際の作業に応用する際にサポートを受けることができるSlack上のワークスペースも紹介している。そして終了時に、講座を評価するために、受講者に講座で良かった点一つと改善が必要な点一つを記入してもらっている。

サービスの評価

筆者らは講座の内容と提供方法を改善することに留意してきた。講座を計画する上では、すでに実施されている授業もしくはイベントの枠内で取り組むことに

大きな利点があった。それは、セミナーやイベントの主催者と協働すれば自分たちで受講者を確保する必要がないこと、そして、講座の前に受講者を具体的に把握できることで例として使用する適切なデータセットを用意しておくことができる点である。一方で、独自に単独の講座を実施した際には、参加者は少なくなり、オンラインで入手できる一般的なデータセットを使わざるを得なかったことがある。

また、可能な限り各々の「データ視覚化ロードショー」を実施してから1週間以内に振り返りを行った。その結果の一例として、講座で用いた全ての視覚化手法を、効果的に情報を伝えつつも、受講者が自分でも作成できると感じさせるくらいのシンプルな内容としたことがあった。これによって、結果として筆者らにとっても、より持続的に発展させることができる内容となった。この決定に至ったのは、土木工学の学生に対して交差点の交通シミュレーションを複雑に視覚化した際の経験による。学生はその複雑さを興味深く感じてはいたが、アニメーションが伝えようと意図した情報を理解するのを難しく感じていたのである。

データ視覚化のベストプラクティスについて説明する部分も改善を必要とした。ここでは元々は、Edward Tufte による効率的で偏りのない視覚化のための考え方⁽²³⁾を提示していた。しかし、不必要な美的効果の使用を最低限にとどめる、といった一般的に望ましいとされる方法は、すでに通常の視覚化ソフトウェアでは標準となっていた。このことから筆者らはこの部分を、視覚化ソフトウェアを利用する・しないに関わらず適用でき、適切かつ理論的なデータ視覚化の枠組みを示している、先述した Tamara Munzner の“Action, Target Pairs”に置き換えた。10回にわたるセッションを通してこういった変更を加え、データ視覚化への入門となる内容を全学にわたって提供できる、柔軟性と再利用性のバランスがとれた内容の講座とそのための所定の手順を完成させることができた。

サービスの導入

ここで紹介した講座は、主要なリソースがあれば他の学術研究機関においても利用することができる。前提として、講座の提供者はデータ視覚化におけるベストプラクティスや利用可能なソフトウェアの全体像を理解している必要がある。研究支援を担当する筆者2人とも、講座の内容の計画と提供、講座における受講者からの質問への対応において、自分達のデータ視覚化の経験を援用することができた。この講座は対面でもオンラインでも提供することができ、受講者が実践的演習に参加するためにはウェブが動作するコン

ピュータと最新のウェブブラウザのみを必要とする。講座の準備に必要な時間は求められるカスタマイズの程度による。Motor Trends 誌の1974年の試乗時燃費のデータセット⁽²⁴⁾や3種のペンギンの計測データ⁽²⁵⁾のような一般に入手できるデータセットを利用する場合は、次回のために講座の内容を変更する労力は必要ない、もしくは少ない。想定される受講者の専門領域からデータセットの事例を得る場合には、適切なデータセットを見つけることを含むカスタマイズ、データセットに適合するコードへの変更、スライドの変更に2時間～4時間の準備時間を必要とする。講座自体は質疑応答とトラブル対応も含んで1時間で計画されている。講座の例は Google Slides から⁽²⁶⁾、プレゼンテーションで使用した Jupyter Notebooks のリストは GitHub から⁽²⁷⁾入手できるようにしている。これら全ての素材には改変と再利用を容易とするために許容的なライセンスを付与している。

※本稿はデヴィン R. バイリー氏とジェフリー C. オリヴァー氏による“The Data Visualization Roadshow: A data visualization session provided by an academic library”の全訳である。原文は以下を参照のこと。

https://current.ndl.go.jp/en/ca2031_en

- (1) National Academies of Sciences, Engineering, and Medicine. Data Science for Undergraduates: Opportunities and Options. National Academies Press, 2018, 138p. <https://doi.org/10.17226/25104>, (accessed 2022-09-16).
- (2) Aghassibake, Negeen et al. Supporting Data Visualization Services in Academic Libraries. The Journal of Interactive Technology and Pedagogy. 2020, 18. <https://jitp.commons.gc.cuny.edu/supporting-data-visualization-services-in-academic-libraries/>, (accessed 2022-09-16).
- (3) Cox, Andrew M. et al. Developments in Research Data Management in Academic Libraries: Towards an Understanding of Research Data Service Maturity. Journal of the Association for Information Science and Technology. 2017, vol. 68, no. 9, p. 2182-2200. <https://doi.org/10.1002/asi.23781>, (accessed 2022-09-16).
- (4) Zakaria, Mahmoud Sherif. Data Visualization as a Research Support Service in Academic Libraries: An Investigation of World-Class Universities. The Journal of Academic Librarianship. 2021, vol. 47, no. 5, 102397. <https://doi.org/10.1016/j.acalib.2021.102397>, (accessed 2022-09-16).
- (5) Aghassibake, Negeen et al. op. cit.
- (6) LaPolla, Fred Willie Zametkin; Rubin, Denis. The “Data Visualization Clinic”: A Library-Led Critique Workshop for Data Visualization. Journal of the Medical Library Association. 2018, vol. 106, no. 4, p. 477-482. <https://doi.org/10.5195/jmla.2018.333>, (accessed 2022-09-16).
- (7) LaPolla, Fred Willie Zametkin. Excel for Data Visualization in Academic Health Sciences Libraries: A Qualitative Case Study. Journal of the Medical Library Association. 2020, vol. 108, no. 1, p. 67-75. <https://doi.org/10.5195/jmla.2020.749>, (accessed 2022-09-16).
- (8) Shao, Gang et al. Exploring Potential Roles of Academic Libraries in Undergraduate Data Science Education Curriculum Development. The Journal of Academic Librarianship. 2021, vol. 47, no. 2, 102320.

- <https://doi.org/10.1016/j.acalib.2021.102320>, (accessed 2022-09-16).
- (9) Oliver, Jeffrey C. et al. Data Science Support at the Academic Library. *Journal of Library Administration*. 2019, vol. 59, no. 3, p. 241-257.
<https://doi.org/10.1080/01930826.2019.1583015>, (accessed 2022-09-16).
 - (10) The R Project for Statistical Computing.
<https://www.r-project.org>, (accessed 2022-09-16).
 - (11) "Python Language Reference". Python.
<https://docs.python.org/3/reference/>, (accessed 2022-09-16).
 - (12) Voyager.
<http://vega.github.io/voyager/>, (accessed 2022-09-16).
Wongsuphasawat, Kanit et al. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Transactions on Visualization and Computer Graphics*. 2016, vol. 22, no. 1, p. 649-658.
<https://doi.org/10.1109/TVCG.2015.2467191>, (accessed 2022-09-16).
 - (13) RawGraphs.
<https://www.rawgraphs.io>, (accessed 2022-09-16).
 - (14) Auriga, Radosław et al. "Data for: Performance Properties of Plywood Composites Reinforced with Carbon Fibers". Mendeley Data. 2020-06-07.
<https://doi.org/10.17632/HKGSTM9SXG.1>, (accessed 2022-09-16).
 - (15) Munzner, Tamara. *Visualization Analysis & Design*. 1st edition, CRC Press, 2015.
 - (16) Field, Kenneth. Thematic Maps of the 2016 Presidential Election (lower 48 states).
<https://carto.maps.arcgis.com/apps/MinimalGallery/index.html?appid=b3d1fe0e8814480993ff5ad8d0c62c32>, (accessed 2022-09-16).
 - (17) Kermack, W. O.; McKendrick, A. G. Contributions to the Mathematical Theory of Epidemics—I. *Bulletin of Mathematical Biology*. 1991, vol. 53, no. 1-2, p. 33-55.
<https://doi.org/10.1007/BF02464423>, (accessed 2022-09-16).
 - (18) Yeh, I. C. Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks. *Cement and Concrete Research*. 1998, vol. 28, no. 12, p. 1797-1808.
[https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3), (accessed 2022-09-16).
 - (19) Project Jupyter et al. "Binder 2.0 - Reproducible, Interactive, Sharable Environments for Science at Scale". *Proceedings of the 17th Python in Science Conference (SciPy 2018)*. 2018, p. 113-120.
<https://doi.org/10.25080/Majora-4af1f417-011>, (accessed 2022-09-16).
Jupyter.
<https://jupyter.org/>, (accessed 2022-09-16).
Pérez, Fernando; Granger, Brian E. IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*. 2007, vol. 9, no. 3, p. 21-29.
<https://doi.org/10.1109/MCSE.2007.53>, (accessed 2022-09-16).
 - (20) Observable.
<https://observablehq.com/>, (accessed 2022-09-16).
Bostock, Michael et al. D³: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*. 2011, vol. 17, no. 12, p. 2301-2309.
<https://doi.org/10.1109/TVCG.2011.185>, (accessed 2022-09-16).
 - (21) "Bar Chart Race, Explained". Observable.
<https://observablehq.com/@d3/bar-chart-race-explained>, (accessed 2022-09-16).
<https://www.rawgraphs.io>, (accessed 2022-09-16).
 - (22) VIS 2021 Satellite Sites.
<http://ieevis.org/year/2021/info/satellite>, (accessed 2022-09-16).
 - (23) Tufte, Edward R. *The Visual Display of Quantitative Information*. 2nd ed, Graphics Press, 2001.
 - (24) Henderson, Harold V.; Velleman, Paul F. Building Multiple Regression Models Interactively. *Biometrics*. 1981, vol. 37, no. 2, p. 391-411.
<https://doi.org/10.2307/2530428>, (accessed 2022-09-16).
 - (25) Horst, Allison M. et al. "Allisonhorst/Palmerpenguins: V0.1.0". Zenodo, 2020.
<https://doi.org/10.5281/ZENODO.3960218>, (accessed 2022-09-16).

- (26) Oliver, Jeff; Carini, Kiri; Bayly, Devin. Data Visualization.
<https://docs.google.com/presentation/d/1g13QTzIWMia30kOzTEJ-LP3GD7iZL-9xgu-geq05bOg/edit?usp=sharing>, (accessed 2022-09-16).
- (27) "jcoliver/data-viz-roadshow". GitHub.
<https://github.com/jcoliver/data-viz-roadshow>, (accessed 2022-09-16).

[受理:2022-11-16]

Devin R. Bayly

Jeffrey C. Oliver

Translation : Kamada Hitoshi

The Data Visualization Roadshow: A data visualization session provided by an academic library



本著作 (CA2031) はクリエイティブ・コモンズ 表示 4.0 国際 パブリック・ライセンスの下に提供されています。ライセンスの内容を知りたい方は <https://creativecommons.org/licenses/by/4.0/legalcode.ja> でご確認ください。