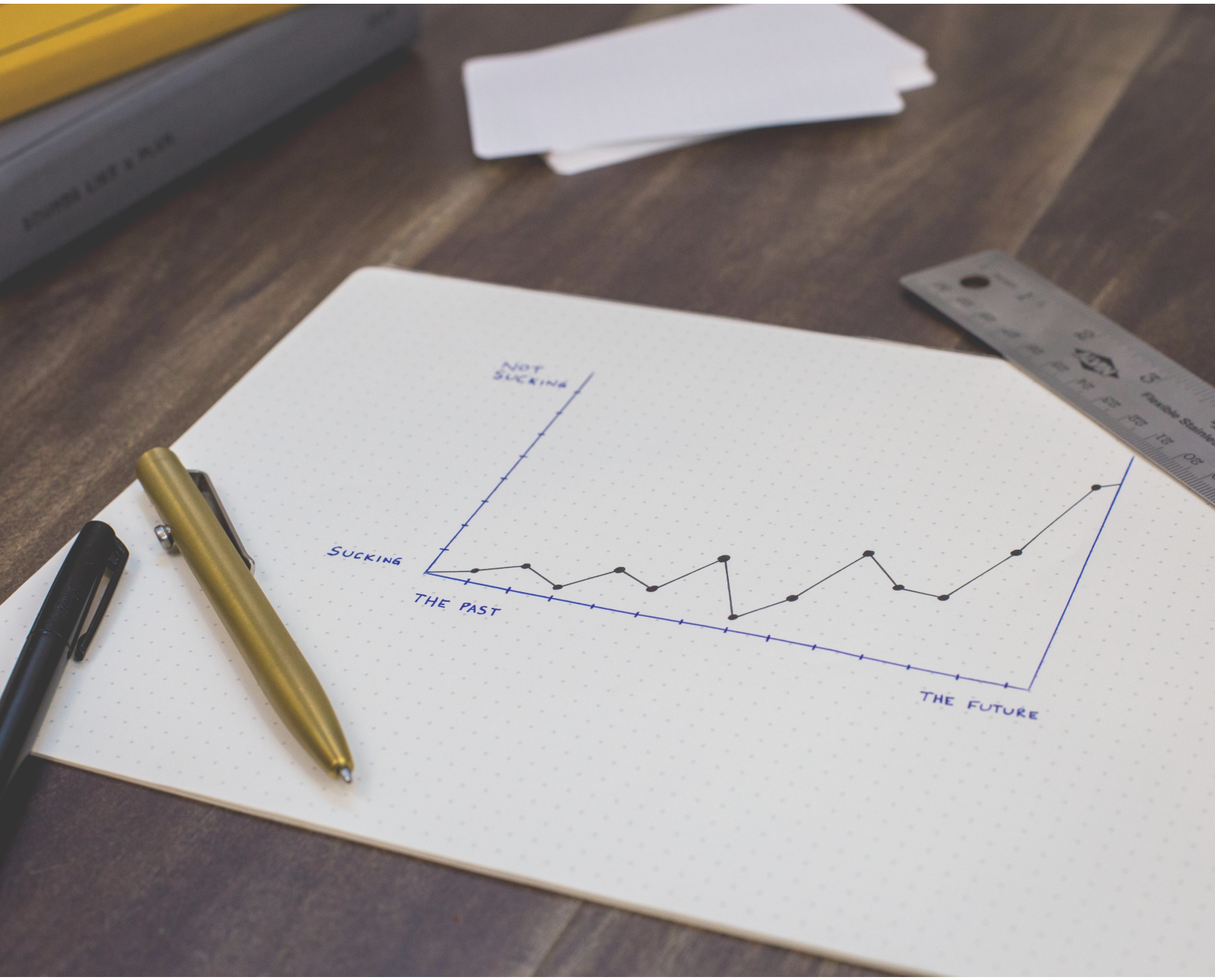
ORMS TOMORROW

informs | SPRING / SUMMER 2022



FOREGROUNDING OPERATIONS RESEARCH
In Modern Art

URBAN FREIGHT AND PARKING ISSUES
Interview with Dr. Abdelrahman Ismael

A SCAFFOLDED ASSIGNMENT IMPLEMENTATION
For a Data Analytics Class

CONTENT

Letter from the Lead Editors 3

Foregrounding Operations Research in Modern Art 4

In the Wild: Operations Research in Ecological Studies 8

A Scaffolded Project Implementation For Teaching Undergraduate Data Analytics Course 10

Animated Linear Projections 17

Urban Freight and Parking Issues: Using OR to Improve the Livability of City Centers 20

Letter from the Lead Editors

Dear OR/MS Community Members,

We are pleased to share with you the Spring/Summer 2022 edition of OR/MS Tomorrow!

We welcome three new members to the team:

- Saeedeh Dehghani, as staff writer
- Nandan Kumar Singh, as production editor
- Sofia Perez-Guzman, as co-lead editor

We also thank all of our departing members for their great service and contributions to our magazine: Srinivasan Balan, Egbe-Etu Emmanuel Etu, Bukola Bakare, Elham Taghizade, Armagan Ozbilge, Saba Moeinizade, Piyal Sarkar, and Amro El-Adle. We wish you all a bright future!

In this issue, we discuss a myriad of topics in different areas, including the role of operations research in non-traditional OR/MS topics of ecological studies and modern art. From there, the issue jumps into data analysis, with an article presenting a scaffolded project implementation to teach data analytics to undergraduate students, and an article discussing how to visualize multidimensional data and use tours to help analysts to glean information from data. Lastly, the issue moves toward transportation matters with an interview with Dr. Ismael, who discusses urban freight and parking issues.

Submissions for the Fall/Winter 2022 issue are welcomed. Contact us via email at orms_tomorrow@mail.informs.org!

We hope you enjoy this issue.

Jessica Leung and **Sofia Perez-Guzman**

OR/MS Tomorrow Editorial Team

This issue wouldn't have been possible without the effort of our editorial team members:

Editorial Staff Writers
Saeedeh Dehghani, Abigail Lindner, Mihir Mehta, Kimia Vahdat

Editorial Board Members
Amro El-Adle, Abtin Ijadi Maghsoodi, Yassine Yaakoubi

Production Editor (Latex)
Nandan Kumar Singh

Webmaster
Harsh Anand

Social Media Coordinators
Kara Combs, Armagan Ozbilge

Faculty Advisor
Prof. David Czerwinski



Foregrounding Operations Research in Modern Art



Abigail Lindner
Worcester Polytechnic Institute



Nandan Kumar Singh
Indian Institute of Management
Visakhapatnam

Introduction

The mathematician does not study pure mathematics because it is useful; he studies it because he delights in it and he delights in it because it is beautiful. - Henri Poincaré (1854-1912)

Operations research (OR) has a long history of contributing to supply chain management, transportation routing, and other business problems. Less well-known is the use of OR tools in non-technical fields. In the humanities, OR appears in the production of pieces of modern art via reframing of the Traveling Salesman Problem and the application of linear optimization and graph partitioning.

Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is a classical question in operations research. The objective in the TSP is to identify the best route through N cities, represented by points, that visits each city exactly once and returns to the origin city. The "best" route might be the shortest (distance), least expensive (sum of the costs of each leg of the journey), or a combination of the two.

Dr. Robert Bosch, James F. Clark Professor of Mathematics at Oberlin College & Conservatory, has gained a reputation for creating "optimization art," which he details in his 2019 book *Opt Art: From Mathematical Optimization to Visual Design* (Bosch, 2019). Among his inventions is visual art made by treating classic paintings and prints as canvases on which a TSP may be overlaid.

Consider the Mona Lisa, by Leonardo da Vinci, arguably the most famous painting in the world. In place of the lines and curves that form the original painting, Bosch converted the Mona Lisa into a series of tens of thousands of dots representing "cities" in the TSP (Cook, 2022). The optimal solution to the TSP is the continuous-line drawing of the original Mona Lisa that is as accurate as possible.

For small-scale routes of, say, 5 cities, an analyst could determine the optimal route by trying each of the $\frac{(5-1)!}{2} = 12$ possible solutions. Dealing manually with the 100,000 "locations" that make up Bosch's Mona Lisa problem, however, would be a feat of



Figure 1: Mona Lisa TSP challenge, From Robert Bosch, William Cook, and Yuichi Nagata [1]

inhuman endurance and an impossible lifespan. Hence, mathematical optimization tools, which are often employed in OR work, are essential to this TSP art (Klotz, 2015).

These experiments in reproducing works of traditional art through mathematical methods are a fun exercise of optimization theory, and worth trying. In the case of the Mona Lisa, successful efforts to find a route shorter than current best solution also entail a cash prize of $1,000 (Cook, 2022)! The current best TSP solution for the Mona Lisa, depicted above, was found in 2009 by Yuichi Nagata and has length 5,757,191, which is 107 more than the lower bound that Cook has identified for an optimal solution (Cook, 2022).

Domino Mossaic

Consider a single set of dominoes, which contains 55 rectangular tiles. Ignoring some "messy details of geometry and orientation, there are more than 1073 ways of placing the 55 tiles" (Hayes, 2020). Now consider 44 sets of dominoes. There are millions of ways to place all 2,420 tiles. What are the odds that one of these arrangements will produce a portrait of, say, former U.S. president Barack Obama?

The idea of recreating pictures and photos with dominoes - creating domino mosaics - originates with Ken Knowlton, a computer graphics pioneer and artist who worked at Bell Labs in the mid-20th century. During his tenure at Bell Labs, Knowlton experimented with the new capabilities of then-modern technology to create art. He made computer-assisted "mosaic art" with unconventional art tools, including dominoes. (A summary of Knowlton's work is accessible at https://www.knowltonmosaics.com/.)

Bosch drew inspiration from Knowlton to create more complex domino mosaics. As with the TSP art, Bosch aimed to produce the likeness of a monochrome image that "most closely matches the pattern of shadow and light in the original" (Hayes, 2020). His predecessor Knowlton used a two-step process: first, he generated a "pattern of empty domino holders" on a canvas; next, he assigned dominoes to the holders according to brightness, a value based on the number of pips on the domino tile (Cambazard et al., 2011). Bosch diverged by transforming the domino mosaic problem into an integer linear programming problem.

On Bosch's Domino Artwork site, beginners can download guides to produce portraits of Abraham Lincoln and Martin Luther King, Jr., or recreations of the Mona Lisa and the Statue of Liberty using 12 sets of dominoes. The downloads page is at http://www.dominoartwork.com/downloads.html.

[1] Source: https://www.math.uwaterloo.ca/tsp/data/ml/monalisa.html
The Mona Lisa TSP was commissioned by William Cook. The point set was designed by Robert Bosch and the tour was obtained by Yuichi Nagata. The image can also be found on page 108 of Robert Bosch's *Opt Art*.

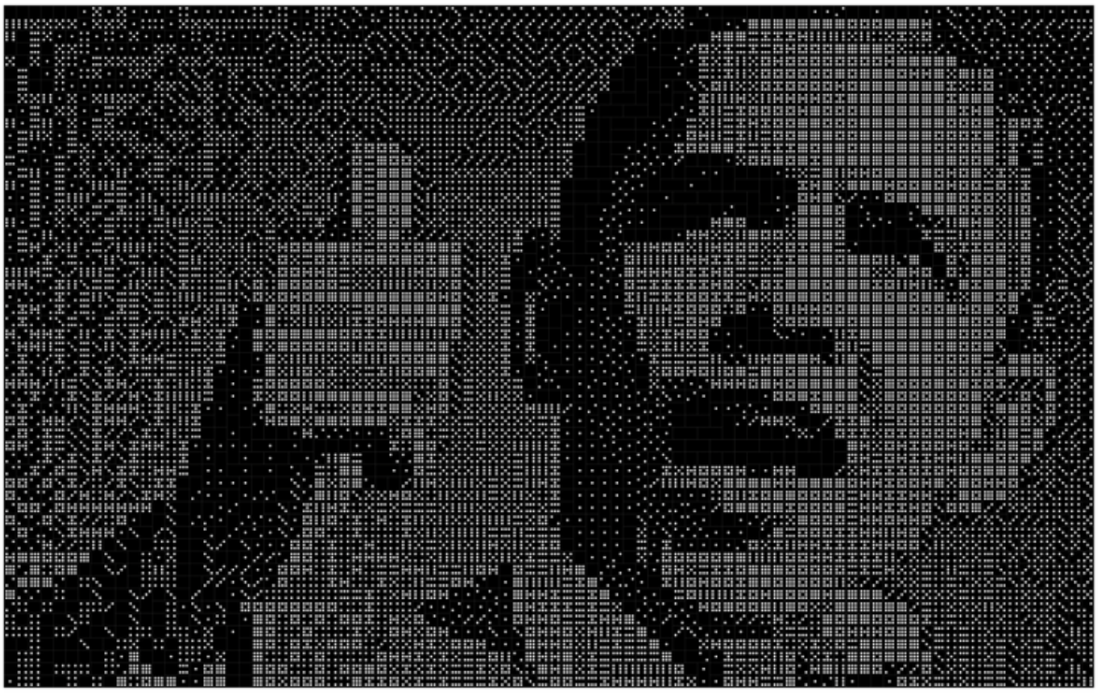


Figure 2: From Robert Bosch, 2008 [2]

Graph Partitioning

Graph theory is a discipline focused on graphs, mathematical structures that model pairwise relationships between objects. It began as a recreational math problem, but it has now evolved into a prominent field of mathematics with applications in chemistry, operations research, social sciences, and computer science.

Partitioning a graph is a critical step in parallelizing graph algorithms ("Graph partitioning"). It divides a complex problem into multiple, more manageable sub-problems that may be solved in parallel. Graph partitioning algorithms are executed using either edge or vertex separators, depending on the algorithm. Effective partitioning can reduce the time and effort to solve the problem.

How does graph partitioning connect to art? Consider INFORM, a software company which provides digital decision making based on Artificial Intelligence (AI) and Operations Research. The company uses an effective algorithm for graph partitioning to solve routine projects, the results of which have surprising artistic value. One context in which an algorithm developed by INFORM was used for a routine project and transformed into art is a bill of materials structure (Keppner, 2015b).

A bill of materials (BOM) details the components and assemblies needed to construct a comprehensive network structure. Oleg Shilovitsky, a software developer and the CEO of OpenBOM, highlights the importance of graph theory in transforming the BOM into a graph that will be helpful in production planning and analysis (Shilovitsky, 2020a). The network artifact that results from the synthesis of many BOMs can serve as a model for analyzing product information and the relationships between the components that comprise the graph (Shilovitsky, 2020b). Shilovitsky notes that this network can also be a source of artistic interest.

For Keppner, the BOM structure above is where he first realized the artist value of supply chain project visualizations. In this, the 5,564 nodes represent a completed product or a precursor to a complete product and nodes of the same color share the same production resource. Greater size indicates that the completed product consumed a higher percentage of the given resource. This graph is an aggregation of 52 mappings of weekly production capacities over a single year.

Unlike the TSP art and domino mosaics, the creators of these graphs had no intention of creating art. Keppner saw this graph displayed on the wall behind a colleague's desk and assumed that it was a work of abstract art. His mathematical colleague clarified the content of the graph, but Keppner still thought, "*This should be framed.*" Unintentionally, the BOM structure mimicked features of modern art, and from that day forward Keppner began looking at his work through a more visual, artistic lens (Keppner, 2015b). Other supply chain projects that use graph partitioning algorithms, such as workforce planning and aircraft scheduling, exhibit similar features.

[2] Source: https://www2.oberlin.edu/math/faculty/bosch/barack-obama-domino-plans.html
Permission to use obtained from Dr. Robert Bosch on May 2, 2022. This image can also be found on page 82 of Robert Bosch's *Opt Art*.

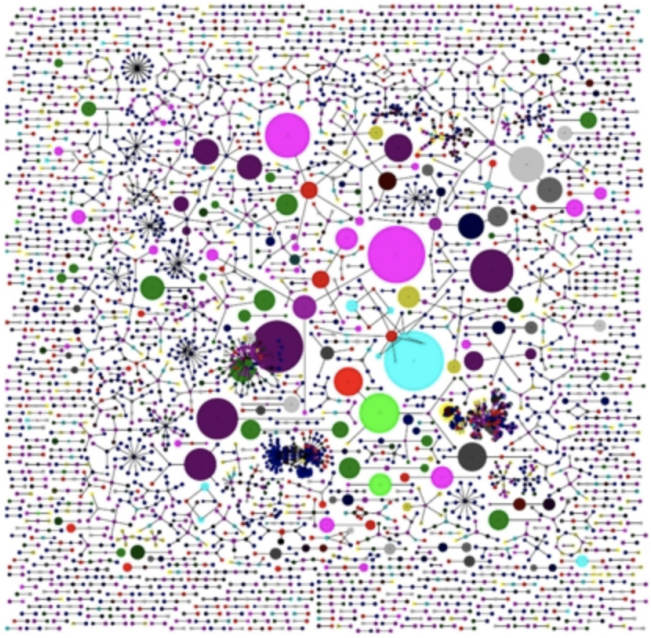


Figure 3: Bill of materials structure [3]

Conclusion

The use of OR in producing modern art is relatively little-known, but there is an amazing store of examples to be discovered. The preceding discussion shares just a small fraction.

Kai Keppner, Head of Corporate Marketing at INFORM, sees two levels of beauty in optimization algorithms translated into graphs: "*the visual form which everyone can appreciate, and the underlying aesthetics of pure numbers which is only accessible to a smaller group of mathematicians*" (Keppner, 2015a). This idea can be applied to the graph partitioning done by decision-making companies and extended to the less conventional graphs of TSP portraits and domino mosaics.

Many thanks to Dr. Robert Bosch for his corrections and inputs on the early drafts of this article.

References

Bosch, R., 2019. Opt Art: From Mathematical Optimization to Visual Design. Princeton University Press.

Cambazard, H., Horan, J., O'Mahony, E., O'Sullivan, B., 2011. Domino portrait generation: a fast and scalable approach. Annals of Operations Research 184, 79–95.

Cook, W., 2022. Monalisa tsp challenge URL: https://www.math.uwaterloo.ca/tsp/data/ml/monalisa.html.

Hayes, B., 2020. Mathematical mosaics. American Scientist. URL: https://www.americanscientist.org/article/mathematical-\mosaics.

Keppner, K., 2015a. Supply chain, operations research & modern art. All Things Supply Chain. URL: https://www.allthingssupplychain.com/supply-chain-operations-research-modern-art/.

Keppner, K., 2015b. The thin line between operations research and modern art. INFORM URL: https://www.inform-software.com/blog/post/the-thin-line-between-operations-research-and-\modern-art.

Klotz, E., 2015. Finding mathematical optimization in unexpected places: The art world. gurobi optimization. INFORM URL: https://www.gurobi.com/resource/finding-mathematical-optimization-in-unexpected-places-\the-art-world/.

Shilovitsky, O., 2020a. Graphs, networks, and boms – part 1. Openbom URL: https://www.openbom.com/blog/graphs-networks-and-boms-part-1.

Shilovitsky, O., 2020b. Openbom: Graphs, networks, and bill of materials – part 3. Openbom URL: https://www.openbom.com/blog/openbom-graphs-networks-and-bill-of-materials-part-3.

[3] Source: https://www.inform-software.com/blog/post/the-thin-line-between-operations-research-and-modern-art
Permission to use obtained from Kai Keppner on May 4, 2022.



In the Wild: Operations Research in Ecological Studies



Abigail Lindner
Worcester Polytechnic Institute

Ecology is a branch of biological science that has existed as a formal field for only about a hundred years, though its history dates much farther back. Practitioners, called ecologists, study the relationships between living organisms and their physical environments to understand the connections between plants, animals, and the surrounding world and illuminate the benefits of ecosystems and humans' influence on them (The Ecological Society, n.d.). Research may be focused on something as small as a single lake teeming with microbes or as large as Earth's species-rich, largely unexplored oceans.

The importance of ecology has been gradually realized in the public and in academia as findings in other branches of biology and other scientific fields spotlight the impact of human activity on modern ecosystems, both natural and man-made, and on the interactions between the living things in them. For instance, the rapid growth of the human population in the last century and the demand for more oil and gas resources have altered the migratory patterns of many large mammals, including bison and elk, in North America, and these changes affect breeding patterns, grassland productivity, and the seasonal dietary habits of some animals (Wilcove, 2010). Experimenting and modeling by researchers in ecological studies can provide insight into what our decisions have affected and how we can respond to the changes in our ecosystems.

Applied mathematics and operations research tools and techniques play an important role in modern ecological studies. In fact, many ecologists come from non-ecology backgrounds, having studied mathematics, statistics, or physics beforehand (Bolker, 2005). These quantitative fields have theories and techniques that provide sufficient supplements and sometimes even substitutes to the empirical and large-scale field experiments that are difficult to replicate or attempt due to the high complexity of environmental conditions (Petrovskii, 2018). Moreover, while the traditional empirical formulae of ecology that use computer simulation techniques are highly adept at fitting curves to data and predicting specific cases, they fail to reveal the underlying ecology. Dynamical systems, often used in operations research logistics problems, can account for these ecological mechanisms ("Mathematical ecology," n.d.).

Historical Background

One of the oldest products of the fusion of applied mathematics and ecology are the Lotka-Voltera models (McCann, 2021). The Lotka-Volterra model, independently designed by American biophysicist Alfred Lotka and Italian mathematician Vito Volterra in the mid-1920s, is the simplest predator-prey model, using a pair of first-order, nonlinear differential equations to describe the dynamics between two species, one predator and one prey.

The "heroic age" of theoretical ecology, before which point ecology was a severely theory-poor biological field, came with the work of Princeton biologist Robert MacArthur, Harvard biologist E.O. Wilson, and their colleagues (Bolker, 2005). MacArthur, in 1965, was "among the first to explore how interactions and competition among species support coexistence and ecosystems," and in 1967 he published the seminal Theory of Island Biogeography with Wilson (Kelly, 2020). This theory supposed

that the rates of migration and extinction in two islands that received species from a common mainland were governed by the distance of each island from the mainland and the size of each island.

Experiments in the latter half of the 20th century, such as one by E.O. Wilson and American biologist Daniel Simberloff involving tiny mangrove islands off the southern coast of Florida, provided ample evidence in support of the theory of island biogeography (Wills, 2013). When physicists began pouring into the field in the 1970s, the use of dynamical systems theory in ecology dramatically increased, enabling the extension of the two-species models of the first half of the 20th century to many more species (McCann, 2021). As a result of these advancements, in the last two or so decades biology has relied heavily on mathematical and computer-based models.

Ecological Research in the 21st Century

Modern ecological research is addressing and exploring a wide variety of problems, including the impact of climate change on marine ecosystems, nitrogen availability in soil affected by agricultural and industrial pollution, the success or failure of conservation efforts to protect underrepresented biomes and endangered species, and the adaptation of wild species to shrinking habitats.

A big topic in ecology is biodiversity, with increased attention and funding being funneled toward research of this topic and increased coordination of biodiversity assessment initiatives (Rillig et al., 2015). Biodiversity refers to the diversity of life on Earth or the "measure of the distribution and abundance of species" (Kelly, 2020). Mathematical models, gene sequencing, climate models, and satellites have expanded our understanding in the last fifty years, though a grand theory unifying all of biodiversity research, which is essential for continuous dialogue between the theoretical and empirical sides of ecology, has yet to be established (Rillig et al., 2015).

On the issue of theoretical ecology and empirical ecology dialoguing and coordinating, a research aim in the 21st century is a greater harmonization of theory and data. Ecology runs into the two extremes of having, in one area, theory without data and having, in another area, data without theory (Kelly, 2020). That is, theories are developed without empirical studies or data is gathered without explicit reference to theory (Rillig et al., 2015).

Dr. Ben Bolker, a mathematical, statistical, and computational biologist at McMaster University and author of *Ecological Models and Data in R* (2008), refers to the more mature scientific field of physics for a good example of the type of relationship he hopes to see in ecology: "empirical and theoretical physicists may distrust each other, but they manage to communicate. Empiricists perform a new experiment, and all the theoreticians go out and try to explain it; theoreticians develop a new theory, and all the empiricists go out and try to test it" (Bolker, 2005).

Possibilities for Operations Research Practitioners

There are numerous examples of operations research practitioners participating in and contributing to ecological research or of operations research tools being used in the same. From modeling invasive species that destabilize ecosystems to identifying optimization methods that take into account efficient resource use and environmental concern to developing ways to minimize the negative impact of human activity on natural ecosystems, operations researchers are applying their quantitative and data-focused skills to address ecological concerns (Mishra, 2020; Büyüktahtakın and Haight, 2018; Dekker et al., 2012).

The skill set of an operations researcher could be easily adapted to the modeling and data demands of modern ecological research. As operations research is primarily concerned with the effectiveness and efficiency of decisions, OR professionals could engage in research on whether and how a given proposal to, say, encourage the repopulation of an endangered species could be executed to maximum benefit with minimal resource expense. Moreover, OR professionals would likely be comfortable poking through the rich store of increasingly publicly available biodiversity data, collecting their own, and making sense of all of it for the benefit of the theoretical ecologists (Rillig et al., 2015).

Opportunities exist on both sides - theoretical and empirical - of ecological research for OR practitioners to make a difference.

References

Bolker, B., 2005. Other people's data. BioScience 55, 550–551.

Büyüktahtakın, I.E., Haight, R.G., 2018. A review of operations research models in invasive species management: state of the art, challenges, and future directions. Annals of Operations Research 271, 357–403.

Dekker, R., Bloemhof, J., Mallidis, I., 2012. Operations research for green logistics–an overview of aspects, issues, contributions and challenges. European journal of operational research 219, 671–679.

Kelly, M., 2020. From muddy boots to mathematics: Advancing the science of ecosystems and biodiversity. Princeton Environmental Institute .

McCann, K.S., 2021. Mathematical ecology. Oxford Bibliographies URL: https://www.oxfordbibliographies.com/view/document/obo-9780199830060/obo-9780199830060-0004.xml.

Mishra, M.K., 2020. Application of operational research in sustainable environmental management and climate change .

Petrovskii, S., 2018. Progress in mathematical ecology.

Rillig, M.C., Kiessling, W., Borsch, T., Gessler, A., Greenwood, A.D., Hofer, H., Joshi, J., Schröder, B., Thonicke, K., Tockner, K., et al., 2015. Biodiversity research: data without theory—theory without data.

Wilcove, D.S., 2010. No way home: the decline of the world's great animal migrations. Island Press.

Wills, C., 2013. Green Equilibrium: The vital balance of humans and nature. Oxford University Press.



A Scaffolded Project Implementation For Teaching Undergraduate Data Analytics Course



Mihir Mehta
Pennsylvania State University

Mihir is a doctoral student at the Industrial and Manufacturing Engineering Department of Penn State University. His research topic is trustworthy artificial intelligence in the public health context. He is an editorial staff writer for the INFORMS OR/MS Tomorrow magazine. He will be a technical mentor for the Data Science for Social Good- UK Chapter Summer Program 2022. He will discuss this article in the "Inclusive Pedagogy" section at the upcoming INFORMS Annual Meeting 2022.

Undergraduate students with little programming and data analytics background often face challenges in learning problem-solving skills. To help all students become better problem-solvers, the author developed a scaffolded project that leverages publicly available datasets. This article describes this scaffolded project implementation. As part of this project, the author created numerous active learning exercises to help students apply data engineering and data visualization-based concepts to provided datasets. Through this engaging and positive learning-focused project, students developed essential data analytics skills useful for areas such as manufacturing, retail, and finance.

This article is organized as follows. First, the project-related key terminologies are introduced. Then, an overview of the underlying project is provided. Subsequently, the instructional process is discussed in detail with an illustrative example, including the prior assessment strategies, the specifics of the scaffolded module, and post-completion activities. Afterward, a list of additional strategies implemented to nurture a positive and engaging classroom experience is provided. The article concludes with a summarized description of the project.

Key Terminologies

- **Instructional Process**: It consists of three stages: *Prior Knowledge Assessment*, *3-section Scaffolded Module*, and *Post Completion Evaluation & Subsequent Development*. This three-stage process develops the specifics to assist with the project completion. It also shapes subsequent course content development. This process is continuously supported through *Complimentary Strategies*. These strategies nurture engaging and positive learning experiences. The entire process is illustrated in figure 4.

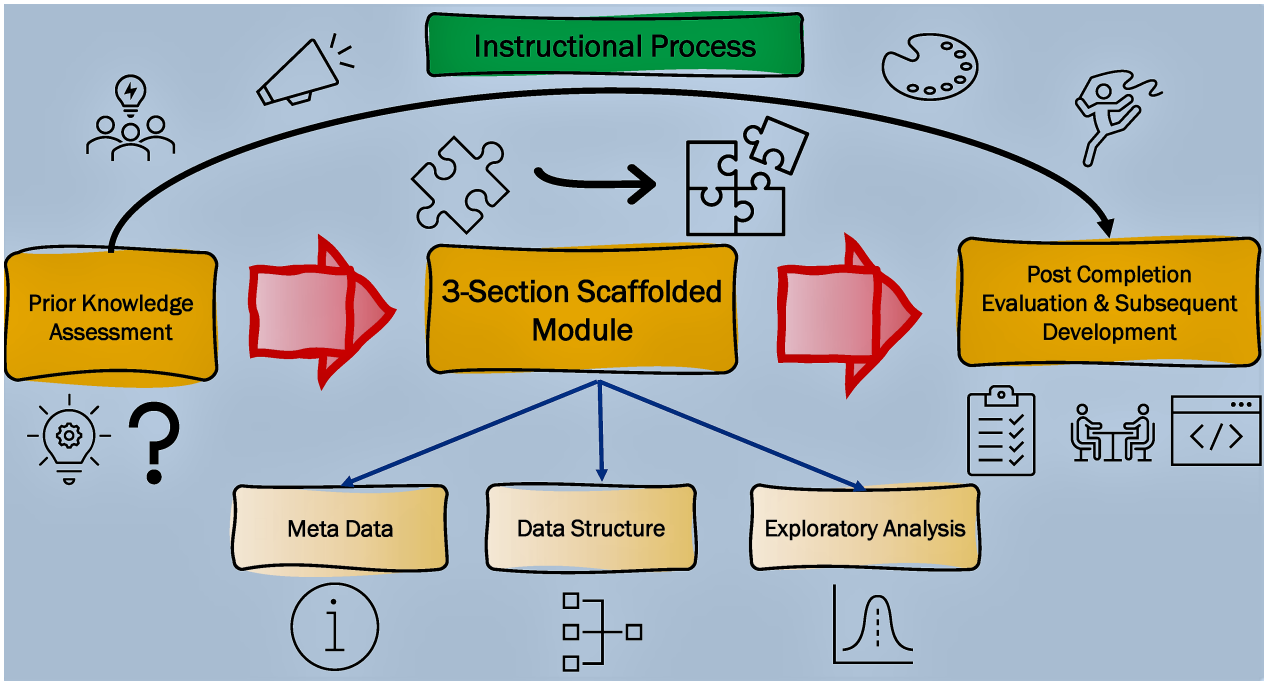


Figure 4: Instructional Process

- **Scaffolded Module:** The 3-section scaffolded module, one of the three stages in the instructional process, helps students to complete this large and complex project through a series of incremental active learning exercises. This module consists of three sections: *Meta Data*, *Data Structure*, and *Exploratory Analysis*. These three sections are motivated by the key learning steps discussed in the project description.
- **Illustrated Example:** It exemplifies the three stages of the *Instructional Process* and the three sections of the *3-section Scaffolded Module*.

Project Description

This section provides a general overview of the implemented exploratory data analytics project. It describes the leveraged publicly available datasets, the format of the problem statements, and the key learning stages in completing project. The next section, instructional process, provides further details illustrated with an example.

Datasets: This project uses three publicly available datasets for the Pennsylvania state:

1. County Level Population Statistics: Dataset (Source Documentation)
2. County Level COVID-19 Cases: Dataset(Source Documentation)
3. County Level COVID-19 Deaths: Dataset (Source Documentation)

Problem Statement: Based on problem descriptions and datasets, students are expected to complete “R" code skeletons to reproduce given set of visualizations.

Key Learning Steps: To reproduce these visualizations, students need to learn incrementally in the following steps:

1. Familiarize with provided datasets (*Meta Data* Section)
2. Understand corresponding data structures (*Data Structure* Section)
3. Formulate data engineering algorithms- a path starting from the provided business problem description to the desired visualization (*Exploratory Analysis* Section)

Instructional Process

This section discusses the proposed instructional process using the project example shown in figure 5, which illustrates the process of going from the project description and dataset to a desired visualization by following the *3-section Scaffolded Module*.

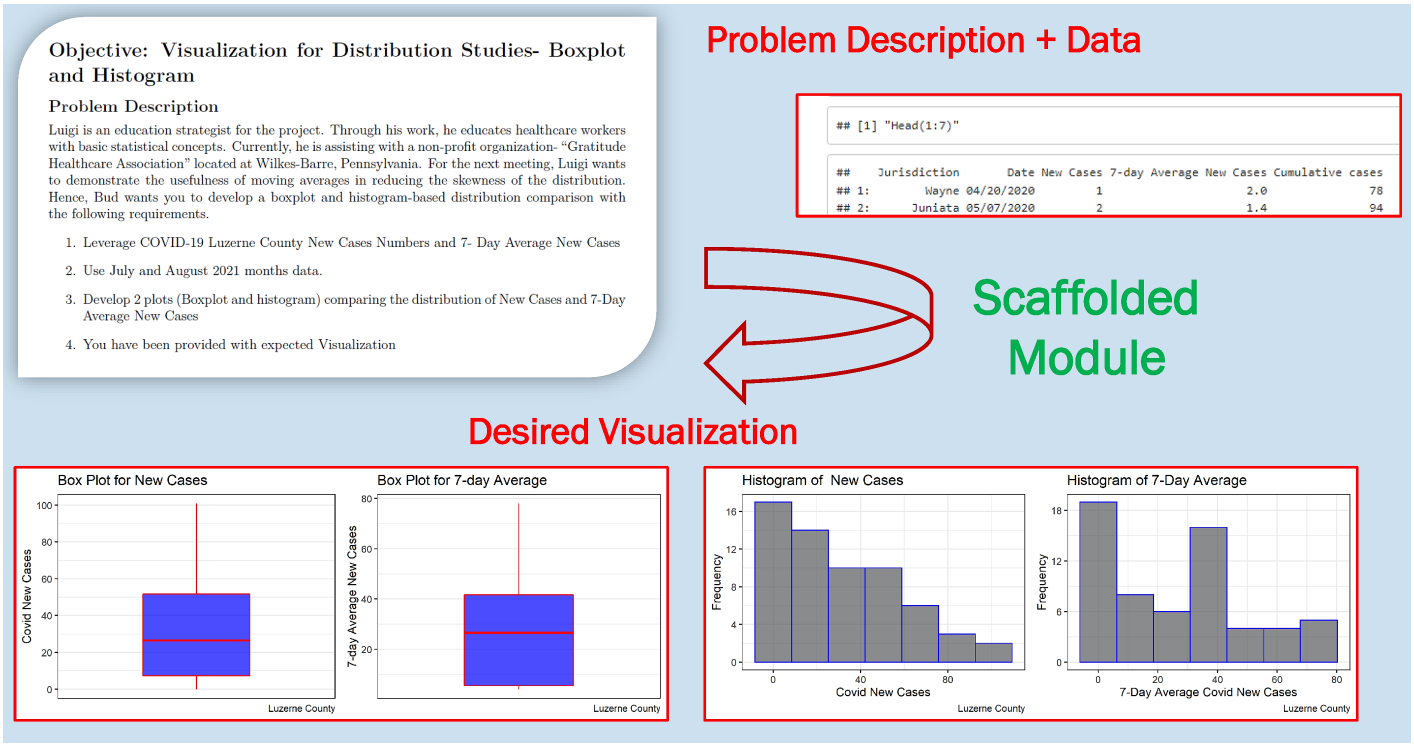


Figure 5: Illustrative Project Example

Using the "County Level COVID-19 Cases" dataset, this example expects the students to (1) do data distribution analyses for the Luzerne County and (2) examine distribution skewness differences between two variables: new case numbers and 7-day average new cases. The project requires reproducing the two visualizations: box-plot and histogram, as shown in figure 5. The *3-section Scaffolded Module* helps the students achieve the goals as follows. The *Meta Data* section helps the students familiarize with different variables in the "County Level COVID-19 Cases" dataset. The *Data Structure* section helps them understand the underlying data structure of the "County Level COVID-19 Cases" dataset. The *Exploratory Analysis* guides them to apply that understanding in reproducing desired visualizations.

Prior Knowledge Assessment

In the first stage of the instructional process, the students' prior knowledge levels are assessed through a set of strategies. These strategies help in designing specifics of the *3-section Scaffolded Module*.

- **Assignments and Classroom Activities:** These activities focus on identifying gaps in programming concepts. **Example:** For a custom-developed R-function, the effect of *"print"* and *"return"* usage on R-output was one of the identified gaps.
- **Multiple-Choice Quiz:** Using source documentation and sample data, this quiz measures students' familiarity with provided datasets. **Example:** True or False Question "Luzerne County has a FIPS code value of 42078."
- **Course Interactions:** Through multi-channel engagement ranging from classroom discussions and office hours to anonymous Canvas surveys-based queries, specific hints, and references for the exploratory analytics section. **Example:** Classroom interactions prompted to include guidelines for referencing a variable name inside ,, in *"dplyr"* syntax.

3-section Scaffolded Module

This module, comprised by three sections, help the students advance towards the completion of their project. Prior knowledge assessments-based insights are used to design the details of this module.

Meta Data

The learning objectives of this section are as follows.

1. Familiarize students with underlying datasets' dimensions and columns
2. Reinforce conceptual gaps identified in the *Assignments and Classroom Activities*
3. Create a smaller dataset to study different data engineering operations through *"dplyr"* verbs

Scaffolded Discussion with Example:

- By defining a simplified data structure involving dataset and information types, a custom R function was developed to create the metadata table.
  Example: The part "b" of figure 6 is a partial output snapshot of the metadata table created for the "County Level COVID-19 Cases" dataset (denoted as *'dt_type=2'*).
- This development applied prior covered programming concepts.
  Example: The part "a" of figure 6 shows some of steps developed in the custom R-function. These incorporate applications of *"dim"*(finding data dimension) and "=="(comparison operation), prior covered concepts . Specifically, **"#Step 9"** is developed to reinforce a conceptual gap example discussed in the *Assignments and Classroom Activities*.
- The custom function included multiple printing statements to follow through with the step-wise development of the final output.
  Example: The part "a" of figure 6 shows *"print"* statement with text description defined before each of the steps.



Figure 6: Meta Data Section

Data Structure

The learning objectives of this section are as follows.

1. Develop a strong understanding of all three datasets' data structure
2. Interpret underlying context captured in a row or a set of rows of the dataset
3. Build self-confidence in R-programming through the practice of low-stake active learning coding exercises

Scaffolded Discussion with Example:

- This section provided a heuristic algorithmic outline for completion. By applying a series of basic *"dplyr"* verbs-based data engineering operations, the outline helps understanding the data structure of a dataset irrespective of source documentation availability. This section also included sample codes and desired outputs creating low-stake active learning opportunities.

Example: Figure 7 shows a partial output of three sub-parts, having the names “The number of FIPS Value by Date," “The number of Date Values by FIPS," and “Dataset for Luzerne 04/13/2020." These results are produced for the “County Level COVID-19 Cases" dataset.

- One sub-part in the outline asks students to understand summary statistics, such as the number of unique values in a specific column and the total number of observations in a group. In the other sub-part, data structure comprehension is strengthened through subsetting the provided dataset for specific values in multiple columns.
  Example: The part “a" of figure 7 lists summary statistics produced at “FIPS Code" and “Date" levels for the “County Level COVID-19 Cases" dataset. These summary statistics-based data structure insights are strengthened by a subsetting this dataset for the “Luzerne" county value for the “04/13/2020" date. The corresponding partial output is shown in the part “b" of figure 7.

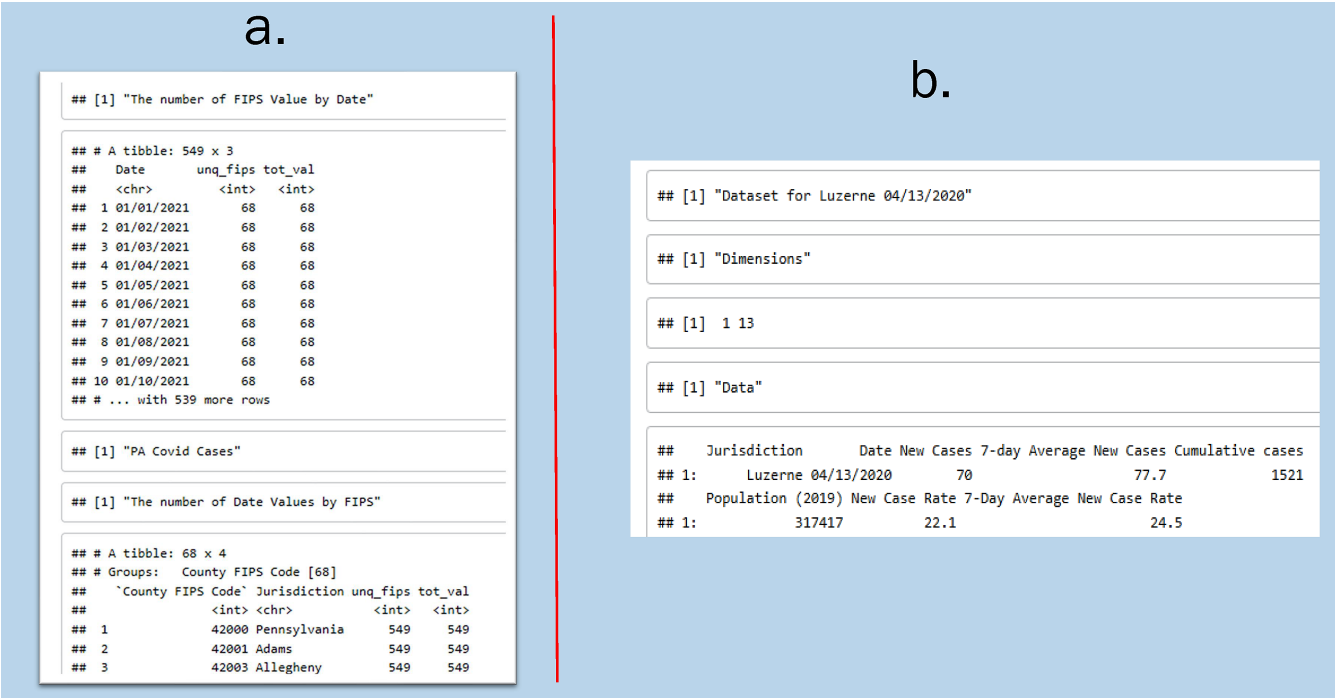


Figure 7: Data Structure Section

Exploratory Analysis

The learning objectives of this section are as follows.

1. Apply data engineering and data visualizations concepts on provided datasets to answer a specific data analytics question using a visualization
2. Provide exposure to the thought process of formulating multiple analytics questions using the same dataset
3. Develop an understanding of the various components of a visualization
4. Familiarize with different practices useful in developing easy-to-interpret visualizations
5. Infer insights about the dataset from a visualization

Scaffolded Discussion with Example:

- This section provided an R-code skeleton to generate desired visualizations.
  Example: The part “c" of figure 8 shows snippets of provided R-code skeleton for distribution studies.
- A set of hints were included to sketch the required algorithm including the mapping for visualization components to *“ggplot2"* specific keywords.
  Example: The part “b" of figure 8 shows some provided guidelines to develop an algorithm for completing distribution studies. The box outline highlights provided *“ggplot2"* mapping.
- A reference HTML document with multiple R-codes and resultant outputs was arranged for illustrating varied applications involving a set of data engineering operations.
  Example: The part “a" of figure 8 contains an R-code and a partial resultant output. This code is producing all the Q1-2021 observations for the “Luzerne" county in the “County Level COVID-19 Deaths" dataset.

- Another reference HTML document was supplied to reinforce some specific concepts identified in the “Course Interactions" assessment.
  Example: The part “d" of figure 8 demonstrates how to reference variables containing spaces in their names. This development is inspired by an interaction discussing “Including variable names inside ,,."

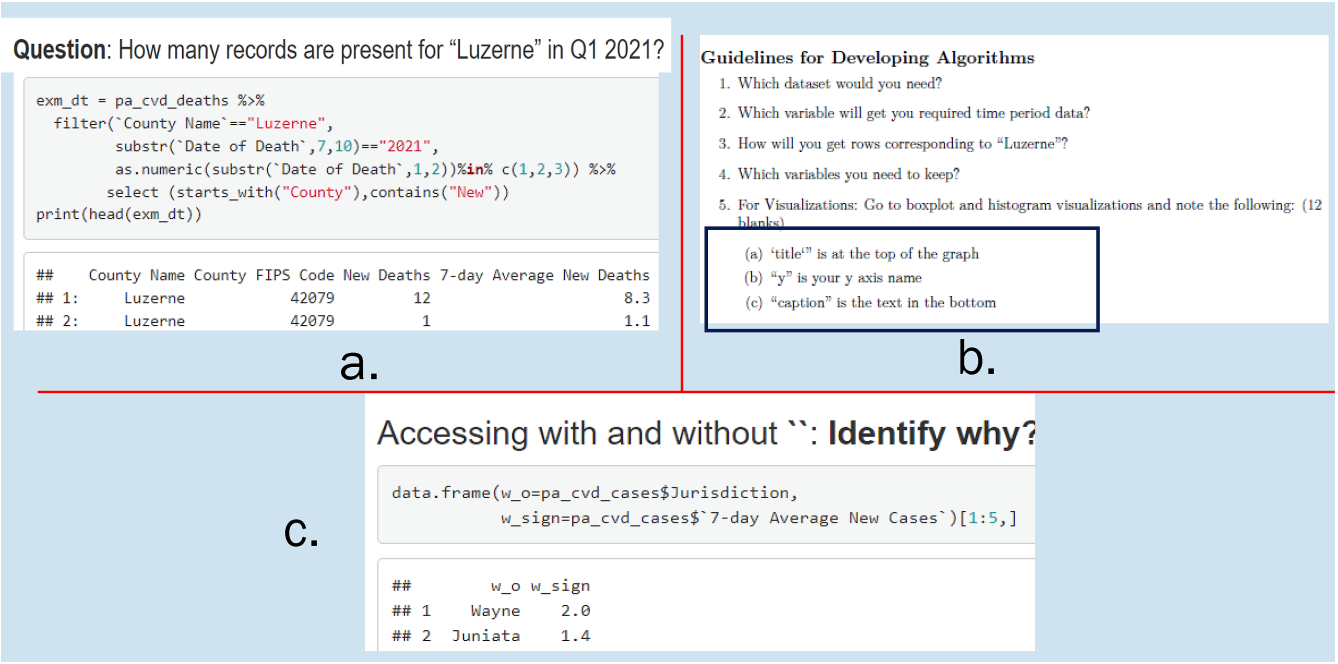


Figure 8: Exploratory Data Analysis Section

Post Completion Evaluation & Subsequent Development

This stage of the instructional process evaluates students’ post-completion comprehension levels. Project evaluations are discussed individually in the 15-minute scheduled oral evaluation. Besides the project feedback, these discussions recommend students a set of strategies for enhancing comprehension levels and learning in the remaining course. Students are strongly encouraged to take an active part in discussions. Students are awarded 30% project grade for participating.

Project evaluations help identify key areas for concept reinforcement. Oral discussions help in deciding content development representation approaches to create an enhanced learning experience. Combined insights from both of these activities shape the design of the subsequent course modules.

Subsequent Content Development with Example:

- **Evaluation:** Parts “a" and “b" of figure 9 show snippets from the content developed for reinforcing data structure and engineering related concepts, respectively. These were identified while evaluating project submissions.
- **Discussion:** The part “a" of figure 9 shows a structural thinking approach designed for learning intricacies of data structure concepts. The part “b" of figure 9 demonstrates a structural format developed to observe and list numerous specifics during data engineering operation. Oral evaluations shaped both of these structural frameworks.
- **Content Example:** Oral evaluations found the *Data Structure* section the most interesting for students. Also, these interest levels were reflected throughout the project evaluation process. Students showed higher comprehension levels for these concepts. Hence, the data governance module and generalized analytics process-related content was designed from the a data structure perspective. The part “c" of figure 9 shows partial outputs of the simulated finance and retail datasets. Data structural similarities between these datasets were highlighted to teach generalized analytics framework and data governance concepts.

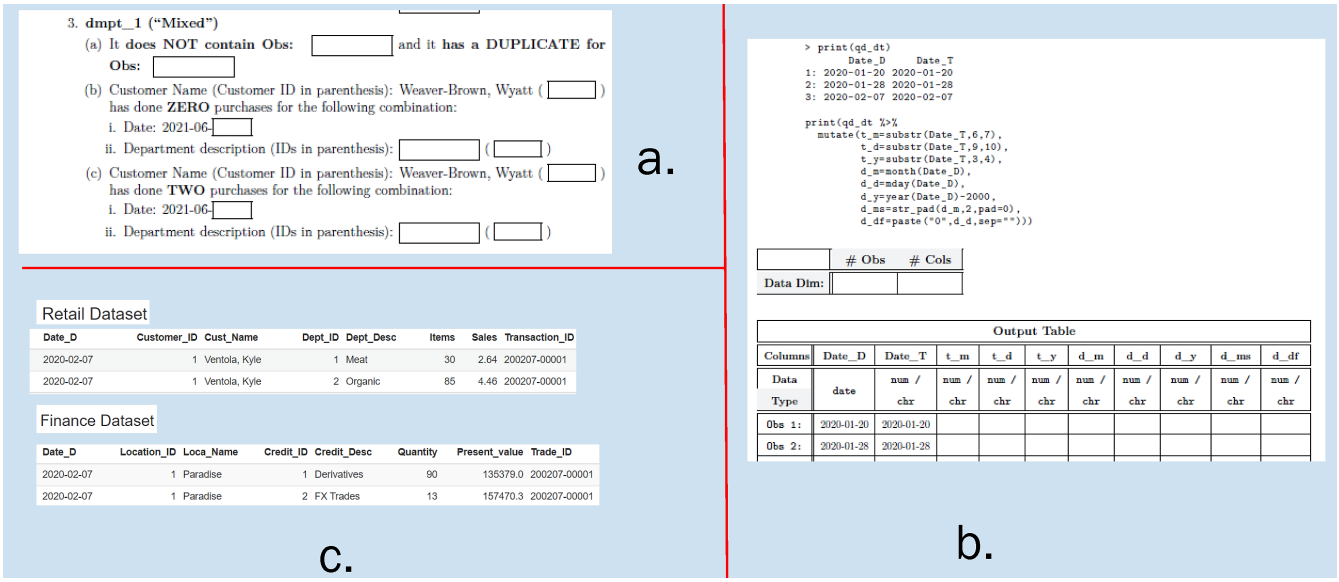


Figure 9: Post Completion Content Development

Complimentary Strategies

The following set of strategies nurtured continuous positive and engaging learning experiences.

- **Communication:** The written communication incorporated a few empathetic and light-hearted lines from time to time. **Example:** "Just like in real life, group dynamics may not always work in your favor. You can still convert those lemons into lemonade. Hence, when you get an Alphonso mango, that smoothie and/or milkshake would be tastier!"
- **Creative Components:** Multiple anonymous creative quizzes were developed on Canvas encouraging divergent thinking. **Example:** Students were asked to describe their fondness for the Penn State football using ice-cream flavors.
- **Flexibility:** Students were provided a few flexible options on submission documents. **Example:** One-page poster submission provided the flexibility for the format and content. Students were encouraged to use this poster submission to express their learning, experiences, and feelings while completing the project.
- **Peer Support:** Students were encouraged to brainstorm in groups. **Example:** Course consisted of a similar exploratory analysis group assignment. A major portion of this group assignment was discussed and solved during course lectures and homework help sessions.

Summary

Data engineering and data visualization are building blocks for solving data analytics problems. This article proposes an exploratory analytics project leveraging publicly available datasets. For that, a three-stage instructional process was developed. After assessing students' prior knowledge levels in the first stage, a *3-section Scaffolded Module* follows as the second stage. This module contains all students' inclusive active learning exercises. In the third stage of the instructional process, there are post-completion assessments and individual discussions with students. Using insights derived from the third stage, a structural framework is incorporated to reinforce some of the key concepts. Also, these insights are used to decide about content representation for subsequent course modules. Finally, the proposed process nurtures a positive and engaging learning environment by executing the complimentary strategies.

Acknowledgment

The author acknowledges the Industrial and Manufacturing Engineering Department of Penn State University, Sofia Perez-Guzman for edits on this article, Larkin Hood from Penn State's Schreyer Institute for Teaching Excellence for her input on the abstract, and Paul Griffin for encouraging the pursuit of the concept.



Animated Linear Projections



Nicholas Spyrison
Department of Human-Centered Computing
Department of Econometrics and Business Statistics
Monash University, Australia

Nick's research focus is on multivariate data visualization, dimension reduction, and nonlinear model interpretability. He is also the author of the R packages 'spinifex' and 'cheem'.

Website: https://nspyrison.netlify.com
Github: https://github.com/nspyrison
Linkedin: https://www.linkedin.com/in/nspyrison/

Introduction

Suppose you are at the doctor's office for a general check-up. How many measures or variables do you think were measured? Certainly age, height, weight, blood pressure, and heart rate — maybe blood oxygenation or iron levels in the case of a blood test. The visualization of quantitative multivariate data quickly becomes difficult as the dimensionality or number of measures increases. The visualization of data conveys much more information than numerical summaries allow and is vital for every field of work. First, let us consider how we can visualize dimensions.

Traditional multivariate visualization

Consider how we visualize data. For single variables, histograms and densities are used to show the spread. Scatterplots can be used to relate information from 2 or 3 measures simultaneously. After that, we have to mix it up. Taking an exhaustive approach, we could consider an arbitrary, p histogram densities. Similarly, we could look at all pairs and triples of scatterplots. This is the crux of scatterplot matrices (Chambers et al., 1983). As the number of dimensions increases, the number of panes produced from this increases exponentially. This also has another limitation; it is not able to show information from more than 2 or 3 dimensions at a time.

Another method is to link each observation with a line and display variables horizontally. Then after normalizing or standardizing each variable to a common scale, they can be viewed as parallel coordinate plots (Ocagne, 1885). This visual will scale linearly with

dimensions but has a few downsides. The amount of ink each observation uses is relatively high, its hard to extract information such as correlation, and leads to asymmetric interpretation when the order of the variables changes.

Projections

As the dimensionality of the data increases, the analyst will want to turn to dimension reduction as projections. This involves a function that maps the data from a larger space to a smaller space. This is separated into two categories, linear and nonlinear. The linear case spans all affine mathematical transformations, essentially any function where parallel lines stay parallel. Nonlinear transformations complement the linear case, think transformations containing exponents or interacting terms.

Intuitively, linear projections can be thought of as 2D shadows of 3D objects. Holding the light and background constant, the shape of the shadow is determined by the object and its orientation. The orientation and map of the data are defined in a matrix called the basis, A where $Y_{n\times d} = X_{n\times p} \times A_{p\times d}$. As a sidenote, we restrict these bases to be orthonormal (columns at right angles and of length 1). One common linear projection is that of principal component analysis that provides a reorientation basis ordered by decreasing variation (found with eigenvector decomposition) (Pearson, 1901). Figure 10 shows two linear projections of penguins data (Gorman et al., 2014). The left frame is informative to the separation of clusters and the right is much less so.

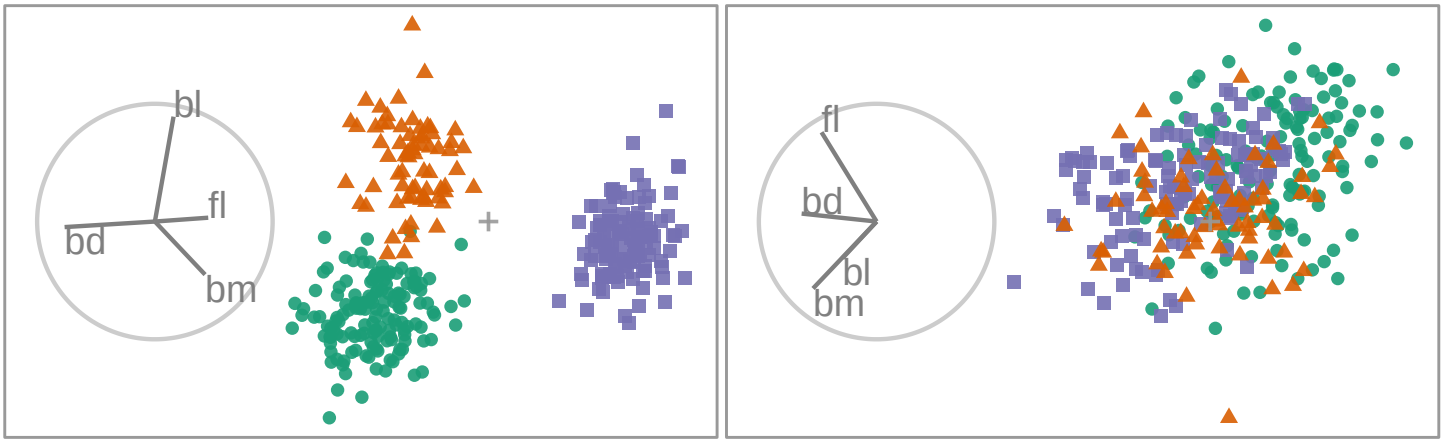


Figure 10: Two basis orientations of penguins data. Some orientations are more informative than others.

Nonlinear projections that can bend and distort spaces are not entirely accurate or faithful to the original variable space. There are various quality metrics, such as Trustworthiness, Continuity, Normalized stress, and Average local error that have been introduced to describe the distortion of the space (Espadoto et al., 2021). Unfortunately these metrics are hard to visualize and communicate, making the distortions introduced opaque to the analyst. The intuition of this can be demonstrated with map projections. Snyder (Snyder, 1987) lists over 200 different projections that distort the surface of the earth to display as a 2D map, each with unique properties and use cases. The added subjectivity of choosing a nonlinear method and its parameters, and the difficulty of understanding how space is distorted makes it hard to interpret variable influences. These factors potentially obscure the signal in the data.

Traditionally linear projections are used to approximate the data in fewer dimensions than the original. This is well and good, but can still leave a gap in visualizing the data. That is, the data approximated down to the first d principal components must be visualized thoroughly rather than just looking at the first 2 or 3 components, as these lower components still regularly hold signal, such as in clustered data (Donnell et al., 1994) or high-throughput genomics data.

Animated linear projections

Going back to the shadow analogy, the analyst can discern information about the shape of the unknown data object by observing the shadow as the object is rotated. In the case of a bar stool, some profiles are relatively uninformative such as a circle from the seat, which could come from many items. However, if the object continually rotates, the legs will show in the shadow, conveying a much better idea of the object's shape. This is the intuition for *tours*. Tours are linear projections animated over small changes to the basis. By viewing the data over differing orientations, the analyst gleans information about the data: which orientations contain

interesting features such as cluster separation or unique profile shapes. Another key aspect is the permanence of observation between nearby frames. That is to say, the analyst can see structure by watching small relative changes between consecutive frames. In contrast, the typical case tries to relate two or more distant bases with no intermittent information conveyed.

There are various types of tours distinguished by the selection of their target bases (Cook et al., 2008). The grand tour animates between randomly selected target frames while a guided tour performs projection pursuit, optimizing for some index function. The manual tour allows for the analyst to manually control the contribution of a selected variable. The manual tour is demonstrated in Figure 11, where the contribution of one variable is being changed which explains most of the separation between the orange and green clusters. Lee et al. (2021) does a great job of discussing recent advances and as they remain a topic of interest. Tours can be produced in the R packages **tourr** (Wickham et al., 2011) and **spinifex** (Spyrison and Cook, 2020).

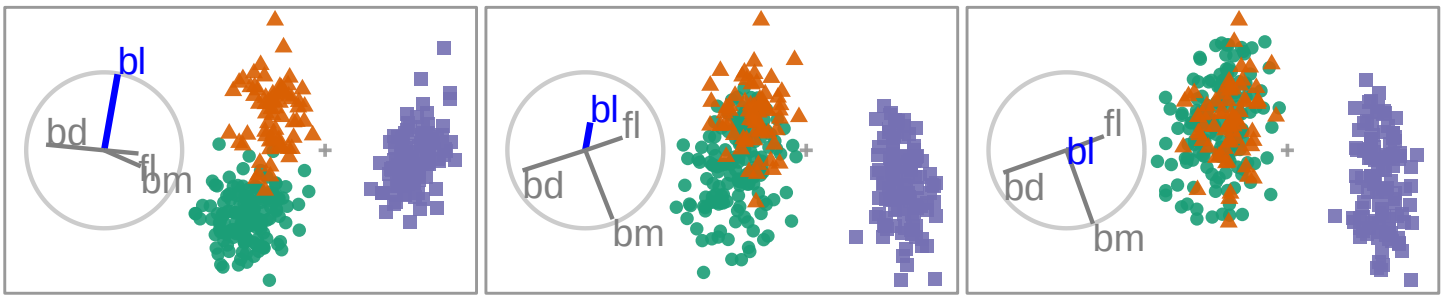


Figure 11: A manual tour varying the contribution of one variable (bill length). As this variable is removed the orange and green clusters overlap. Because of this, we say that this cluster separation is sensitive to this variable. The animated version can be viewed at https://vimeo.com/676723431.

In summary, visualizing quantitative multivariate data remains a difficult task as the dimensions increase. Projections must be used at some point. Traditionally spaces are approximated in fewer dimensions, though these are not often viewed sufficiently. Tours, animated linear projections, can reveal more information than static linear projections by looking any more orientations the data as the basis changes. Recent articles and R packages further discuss and facilitate the production of tours.

References:

Chambers, J., Cleveland, W., Kleiner, B., Tukey, P., 1983. Graphical Methods for Data Analysis .

Cook, D., Buja, A., Lee, E.K., Wickham, H., 2008. Grand Tours, Projection Pursuit Guided Tours, and Manual Controls, in: Handbook of Data Visualization. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 295–314. URL: http://link.springer.com/10.1007/978-3-540-33037-0_13, doi:10.1007/978-3-540-33037-0_13.

Donnell, D.J., Buja, A., Stuetzle, W., 1994. Analysis of Additive Dependencies and Concurvities Using Smallest Additive Principal Components. The Annals of Statistics 22, 1635 – 1668. URL: https://doi.org/10.1214/aos/1176325746.

Espadoto, M., Martins, R.M., Kerren, A., Hirata, N.S.T., Telea, A.C., 2021. Toward a Quantitative Survey of Dimension Reduction Techniques. IEEE Transactions on Visualization and Computer Graphics 27, 2153–2173. doi:10.1109/TVCG.2019.2944182.

Gorman, K.B., Williams, T.D., Fraser, W.R., 2014. Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus Pygoscelis). PloS one 9, e90081.

Lee, S., Cook, D., da Silva, N., Laa, U., Spyrison, N., Wang, E., Zhang, H.S., 2021. The state-of-the-art on tours for dynamic visualization of high-dimensional data. WIREs Computational Statistics , e1573URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1573, doi:10.1002/wics.1573.

Ocagne, M.d., 1885. Coordonne'es paralle'les et axiales. Me'thode de transformation ge'ome'trique et proce'de' nouveau de calcul graphique de'duits de la conside'ration des coordonne'es paralle'les. Gauthier-Villars, Paris.

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2, 559–572.

Snyder, J.P., 1987. Map projections–A working manual. volume 1395. US Government Printing Office.

Spyrison, N., Cook, D., 2020. spinifex: an R Package for Creating a Manual Tour of Low-dimensional Projections of Multivariate Data. The R Journal 12, 243. URL: https://journal.r-project.org/archive/2020/RJ-2020-027/index.html, doi:10.32614/RJ-2020-027.

Wickham, H., Cook, D., Hofmann, H., Buja, A., 2011. **tourr** : An *R* Package for Exploring Multivariate Data with Projections. Journal of Statistical Software 40. URL: http://www.jstatsoft.org/v40/i02/, doi:10.18637/jss.v040.i02.

Interview with Dr. Abdelrahman Ismael

Urban Freight and Parking Issues: Using OR to Improve the Livability of City Centers



Sofia Perez-Guzman
Rensselaer Polytechnic Institute



After earning his B.Sc. in Civil Engineering from Alexandria University, Egypt, and his Master's degree in Transportation Engineering from Rensselaer Polytechnic Institute, Dr. Abdelrahman Ismael completed his Ph.D. at Rensselaer. His main research interests lie in freight transportation and sustainable urban logistics to make deliveries and parking more efficient and to study how to assign vehicles to specific parking spots. The goal of this research is to reduce traffic congestion, parking emissions and the effects of illegal parking.

Transportation is an activity that touches each of us in so many different ways, from using public transit or driving to commute to work and running errands to accessing goods and services in stores or even having them delivered to your doorstep. Regardless of what is transported, transportation is the mere reflection of economic activity, and it is essential to our societies. Nonetheless, despite its vital role as a pillar of our communities, the negative aspects of transportation–such as the negative externalities produced by freight vehicles making urban deliveries–usually receive the most attention. Fortunately, the transportation community has become aware of these negative aspects and has begun to explore alternative solutions.

We invited Dr. Abdelrahman Ismael for an interview to discuss his approach to minimizing urban freight externalities. This article features Dr. Ismael's doctoral research on freight transportation and sustainable urban logistics, conducted at the Center for Infrastructure, Transportation and the Environment at Rensselaer Polytechnic Institute. Dr. Ismael aims to improve the efficiency of delivery and parking activities in urban areas. His research aims not only to improve urban freight operations, but also to significantly reduce the effects of illegal parking, such as traffic congestion and emissions.

Starting with the basics, what is urban freight, why is it important to our society, and how does it relate to parking issues and the livability of city centers?

Urban freight transportation deals mainly with transporting goods in urban areas, whether business-to-business or business-to-consumer goods. The importance of urban freight stems from the congestion and pollution concerns in major cities. Our major urban areas are densely populated, congested and suffer from increased emissions due to many factors. A big one of these is traffic and urban freight vehicles, i.e., trucks. Recently, urban freight issues have received increased attention due to the ever-increasing population of urban areas - 55% of the world's population and 70% projected by 2050 (The World Bank, 2020) – and the concomitant very sharp rise of e-commerce deliveries.

Furthermore, urban freight leads to parking issues because trucks need a place to park to deliver their goods. Sometimes this parking requires additional space for loading and unloading purposes. Not to mention, trucks have larger dimensions, requiring more space per vehicle than passenger cars. Also, trucking activity incurs further difficulties. Truck drivers are usually constrained by having to: (1) meet very tight delivery windows that put them under a lot of rush and pressure; and (2) park in front of their receiver due to specific loading or unloading processes. These two constraints often lead trucks to park illegally. In fact, trucks pay a tremendous amount of parking fines in large cities because they have to deliver their goods, even if they resort to illegal parking. For example, in New York City, two major freight carriers paid around $32 million in parking violations in 2019 (Chang, 2020).

To have further insight, can you briefly tell us what has already been done in this area and what did you find that was lacking?

In general, there is a lot of research on how to manage on-street parking. Most of this research focuses on using dynamic (reactive or predictive) parking pricing, increasing parking spaces to meet demand, using parking reservation systems, and optimally dividing parking spaces among different road users.

One of the major missing factors in this area of research was the accounting of individual vehicles of different vehicle types. Each vehicle may have different characteristics that vary widely, even within the same vehicle type. There is also a problem of lack of information on parking availability. In cities, most vehicle-cruising (when vehicles drive around looking for an available parking space) occurs because we simply do not know where the vacant parking spaces are around our destinations.

How did your work help fill this gap, and what were the main findings from your research?

My work has focused on building optimization models that cities can use to allocate incoming vehicles to different parking spaces, considering their individual characteristics while trying to reduce the overall cost in the network (i.e., achieving a social optimum). In this regard, I proposed two different models (static and dynamic) to deal with the problem.

The main findings were that, generally speaking, vehicles with a higher time value, such as those carrying more valuable goods, have an incentive to park right next to the entrance to their destination. Providing drivers with information about the exact parking location would eliminate cruising altogether, which previous studies have shown to account for an average of 34% of traffic in cities (Hampshire and Shoup, 2019). The net benefit of eliminating cruising can be substantial in major cities that suffer from an even higher percentage of cruising vehicles. Finally, the breakdown in the parking situation as demand exceeds supply does not occur right away, but rather depends on a variety of factors, including changes in demand, parking turnover, and parking duration.

Can you provide an example of how your research can be implemented to have more livable smarter city centers?

The clearest example would be to use this model in a smart city with an infrastructure connectivity system. In such a scenario, drivers would be directed by their navigation apps to a specific parking location decided by a centralized parking allocation system. This system would use driver input, camera/detector information and historical data to make its allocation decisions.

What would be your advice for students who wish to help build smart cities and improve our transportation systems, and what role do OR/MS tools play?

Building smart cities requires a vast amount of data, modeling, and optimization to seamlessly develop control systems that lead to efficient and sustainable transport operations. We all need to get well versed in the modeling aspect, even if the models we produce may not be practical when scaled up. However, these mathematical models can give us good insights into what to do and what not to do. After developing such exact models, we need to be flexible enough to try different paradigms, models, and techniques that may be more practical (e.g., heuristics or trained machine learning models) or offer entirely different perspectives (e.g., bottom-up approaches such as agent-based models). In essence, we need to be flexible in our modeling approaches and perspectives because, ultimately, to quote George Box, "all models are wrong, but some are useful."

Lastly, we need to remember that we are doing this for the welfare and development of humanity. We need to to ask ourselves whether what we are proposing would be in the general interest of our society or would instead increase inequity. We took up science to make the Earth a better place. We should therefore always ponder on the ethical consequences of the products of our research.

References

Chang, S., 2020. Ups, fedex rack up parking violations as city struggles to reduce congestion. https://gothamist.com/news/ups-fedex-rack-parking-violations-city-struggles-reduce-congestion. Gothamist.

Hampshire, R., Shoup, D., 2019. How much traffic is cruising for parking? https://transfersmagazine.org/magazine-article/issue-4/how-much-traffic-is-cruising-for-parking/. Transfers Magazine.

The World Bank, 2020. Urban development. https://www.worldbank.org/en/topic/urbandevelopment/overview.