Let the Data Speak for Itself: Developing a New Data Dashboard for a Hanoi Transport Survey

Kristina Bratkova[*1], Alexis Comber[†2], Phe Hoang Huu[3], Minh Kieu[4], Nick Malleson[‡1,2], Thanh Bui Quang[5], Hang Nguyen Thi Thuy[6], and Eric Wanjau[§1]

[1]Leeds Institute for Data Analytics, University of Leeds, UK
[2]School of Geography, University of Leeds, UK
[3]R&D Consultants, Hanoi City, Vietnam
[4]Faculty of Engineering, University of Auckland, New Zealand
[5]Faculty of Geography, VNU University of Science, Hanoi, Vietnam
[6]VNU Vietnam Japan University, Vietnam National University, Hanoi

January 17, 2022

Summary

Hanoi is a rapidly expanding city with sprawling infrastructure dominated by private vehicles—predominantly motorbikes—that often follow an informal set of traffic rules. Therefore, a bespoke household survey is currently being conducted that collects information about demographics, trip data, vehicle ownership, attitudes towards a proposed motorbike ban and others. This project develops a data dashboard using R Shiny to serve as a communication tool of the survey findings. The interim development of the dashboard is discussed with the breakdown of both the functionality (what the user sees) and the code (the development process and the structure of the app).

KEYWORDS: urban analytics, R Shiny, transport policy, Hanoi

1. Hanoi Transport Survey

The Hanoi population has expanded in the past 50 years, currently standing at 8 million people. As a result, its infrastructure is failing to keep pace with the rapid urbanisation. Moreover, transport is dominated by private vehicles: 89% of Hanoi households own a motorbike or an e-bike and 12% own a car (General Statistics Office, 2019). The biggest problems are noise pollution, air pollution and traffic congestion. Furthermore, as Hanoi population is becomming more affluent, the transition from motorbike to car ownership is likely to further congest the roads (Hansen, 2016).

The transport survey is conducted for the *Urban Transport Modelling for Sustainable Well-Being in Hanoi* project[**] and includes information about the types of journeys people make (reason, distance, travel time), how they make them (transport mode) and for what reason. The examination of these is supported by respondent socio-economic characteristics (age, gender, occupation, etc.), opinions on the motorbike ban, distance to the nearest stop of public transport, and hypothetical questions about alternative modes of transport and potential vehicle ownership. This paper discusses interim progress on the data dashboard, summarising the survey findings that will be used by the policy makers to explore what impact potential scenarios might have. These include a motorbike ban in central districts, increase in car ownership and different distributions of trip-makers among the modes of transport.

Deleted: getting

Deleted: e

Deleted: will

[*] K.Bratkova@leeds.ac.uk
[†] A.Comber@leeds.ac.uk
[‡] N.S.Malleson@leeds.ac.uk
[§] E.Wanjau@leeds.ac.uk
[**] https://urban-analytics.github.io/UTM-Hanoi/

2. R Shiny Dashboard and Geospatial Data

Shiny is an R package used as a framework for creating web applications (Wickham, 2021). Shiny uses reactive programming to track variable inputs and automatically updates related outputs when an input changes. At present, Shiny is relatively simple to pick-up and contains pre-defined configurable layouts, which are visually appealing in their default setting. To illustrate, a dashboard layout relies on relevant CSS and JavaScript, hence the user can focus on the functionality. The dashboard utilises a combination of text content, graphs and maps to communicate the survey findings. Although graphs usually do not pinpoint spatial trends, they are a useful communicator of the data, especially while getting familiar and understanding the trends within the data.

The Shiny application code consists of two fundamental parts: user interface (what the user sees) and server (how Shiny communicates with the widgets to produce the graphics). The user interface (UI) of the application (**Figure 1**) is organised into tabs menus that can be closed and opened to allow the content to occupy full screen width as required.

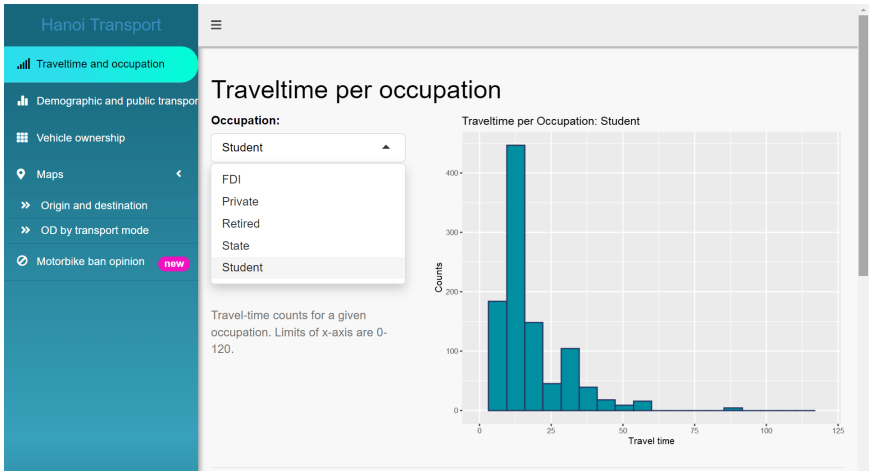

Figure 1 Screenshot of the opening dashboard outlook.

2.1. Filtering and EDA

Filtering is an effective exploratory data analysis (EDA) method that can be used to familiarise oneself with the dataset as well as spot trends. To illustrate, a contingency table of different transport modes and trip frequency can be used to explore peoples' opinions on the motorbike ban (**Figure 2**). The categorical responses on the motorbike ban opinion are converted to integers on a linear scale from 1 (strongly agree) to 0 (strongly disagree). The contingency plot table shows the average for each tile with the widgets on the right, which can be used to update the table in real time. The user can choose 3 different trip frequencies on the x-axis and 3 different y-axis options with filtering on trip numbers are included. The screenshot suggests that the motorbike ban is most favourable among car users and least favourable among motorbike users. As a result, banning motorbikes may implicitly promote cars, which could further exacerbate the traffic congestion as occurred when a motorbike ban was introduced in some Chinese cities (Gallagher, 2006). Unfortunately, the current sample size of approximately 1500 is too small to draw any conclusions. Some tiles have low or even zero counts and further results of **Figure 2** may only be discussed with larger survey sample.

Deleted: used to navigate the page

Deleted: The uppermost tab presents travel time histogram for the selected occupation from the drop-down menu to the left of the plot. The content of Vehicle ownership, Maps: OD by transport mode and Motorbike ban opinion tabs is discussed below.

Deleted: way of

Deleted: , which

Deleted: in certain groups of people

Deleted: the views

Deleted: character

Deleted: of which

Commented [NM1]: I thought the next bit was a little more detail than the reader needs

Deleted: ¶
¶
From the code point of view, the radio buttons used in **Figure 2** return a character string of the selected value, which can be passed onto server part with a unique input variable name of the widget.

Deleted: The input variables are given as x and y axes of the plot and used to filter the relevant data using piping syntax. …

Deleted: it happened

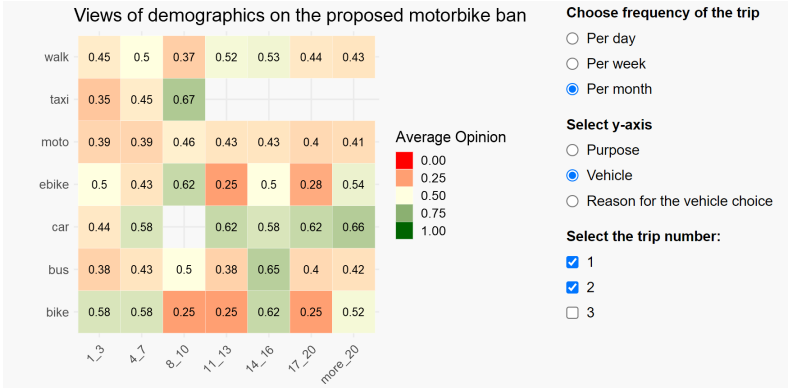

Figure 2 Contingency table (screenshot) of the trip frequencies per month and the vehicle used for the selected trip number (1 and 2).

Figure 3 demonstrates the use of multiple widgets working together. The data a filtered by the vehicle type, number of vehicles considered and the limits of distance to public transport. The plot suggests that the distance to public transport not related to the number of motorbikes owned per household, possibly with the exception of households that own 5 or more motorbikes. As a result, the distance to public transport does seem to be an important factor effecting motorbike ownership, although a larger number of survey responses are required before we can make firm conclusions.

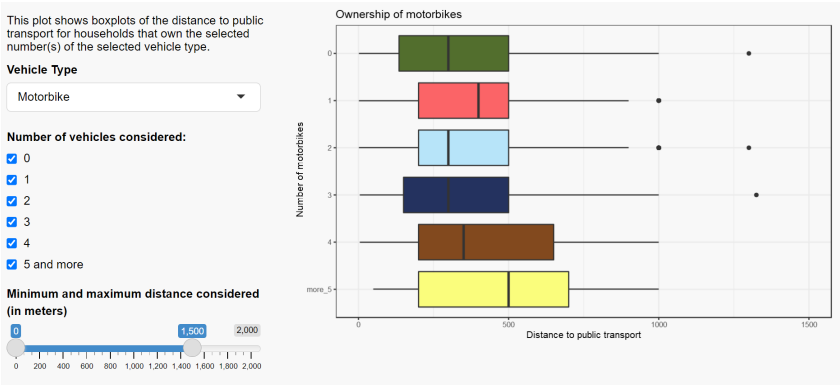

Figure 3 Boxplots (screenshot) of distance to public transport for different numbers of vehicles owned.

2.2. Geospatial Data

A similar filtering approach can be used for mapping origins and destinations of the trips recorded in the survey. **Figure 4** shows the count of origins and destinations per polygon commune of the selected vehicle. Changing the radio buttons prompts Shiny to recalculate the polygon counts instantaneously.

This framework will be used to apply a transport model to investigate the impact of potential policies

Deleted: is

Deleted: The server part of the code uses piping syntax, in which the order of filtering steps is arbitrary.

Deleted: S

Deleted: e

Deleted: of

Deleted: widget

Deleted: when the application is running

such as the motorbike ban. If motorbikes are banned in certain areas or even entire communes, significant traffic redistribution is likely to happen. Guo et al. (2020) analyses the impact of the motorbike ban in Foshan City, China, and concludes that motorbike users are likely to switch to other modes of transport such as cars, e-bikes and public transport or even stop travelling to the area altogether if detrimental implications of the ban are too high. The possible mode transfers could be examined using numeric input from the widgets diverting a certain percentage of motorbike traffic and using the alternative mode of transport specified in the survey to explore possible 'what-if' scenarios.

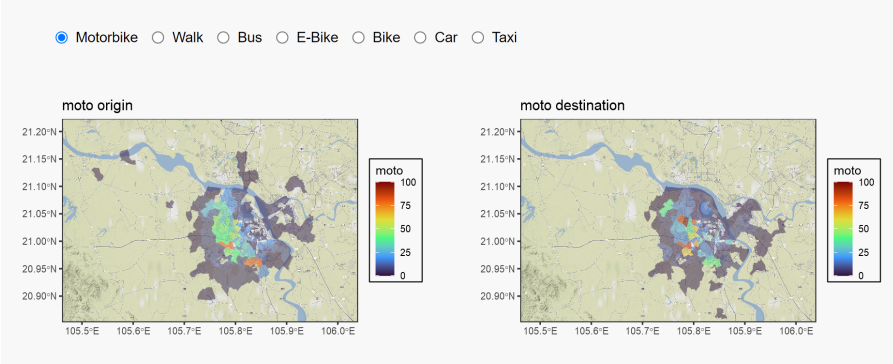

Figure 4 Choropleth maps (screenshot) showing total count of motorbike trip origins (left) and destinations (right) per commune.

3. Conclusion

Extensive functionality specific to R Shiny package includes widgets that can be used in combination to filter data or select parameters and visually appealing UI. Furthermore, shinyapps.io is a server for hosting applications, which can then be shared as a URL link. R Shiny dashboard is an excellent front-end tool to communicate findings of the survey to the policy makers, who can interact with the data without any programming experience. The dashboard will allow policy makers in Vietnam to interrogate the data and take transport modelling input from the project research group to make informed decisions. As well as generating summaries, the dashboard will allow policy makers to explore 'what if' scenarios around attitudes towards a proposed motorbike ban in Hanoi and possible future transport preferences.

Further work seeks to bring the dashboard to production which will include structuring the application as an R package to ensure reproducibility (Fay et al., 2021) and applying the findings to bigger sample. The dashboard is developed to be applied to tens of thousands of survey responses upscaled to the full population of Hanoi using census data and a *propensity score matching* technique. The current survey sample is limited to 1500 responses, thence the research questions may only be hypothesised about. The dashboard tools will be applied to larger samples as the survey collection progresses. Finished work will be published on GitHub along the tenets of open science and reproducible research.

Deleted: performed by Nick Malleson

4. Acknowledgements

This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190].

Special thanks of gratitude goes to Prof Nick Malleson and Prof Alexis Comber that patiently supervised this project and provided valuable knowledge and suggestions.

Deleted: 42

References

Fay, C., Rochette, S., Guyader, V. and Girard, C. 2021. *Engineering Production-Grade Shiny Apps*. CRC Press.

Gallagher, K. 2006. *China Shifts Gears: Automakers, Oil, Pollution, and Development*. Cambridge, MA: MIT Press.

General Statistics Office. 2019. Completed Results of the 2019 Viet Nam and Population and Housing Census [https://www.gso.gov.vn/wp-content/uploads/2019/12/Ket-qua-toan-bo-Tong-dieu-tra-dan-so-va-nha-o-2019.pdf].

Guo, Y., Wang, J., Peeta, S. and Anastasopoulos, P. 2020. Personal and Societal Impacts of Motorcycle Ban Policy on Motorcyclists' Home-to-work Morning Commute in China. *Travel Behaviour and Society*. 19, pp. 137-150.

Wickham, H. 2021. *Mastering Shiny*. USA: O'Reilly Media.

Hansen, A. 2016. Hanoi on wheels: Emerging automobility in the land of the motorbike, *Mobilities*, pp.1-18.

Harvey, F. and Tulloch, D. 2006. Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science*. 20(7), pp.743-768.

Malleson, N., Nguyen Thi Thuy, H., Bui Quang, T., Kieu, M., Hoang Huu, P. and Comber, A. 2021. Urban Data Science for Sustainable Transport Policies in Emerging Economies. *UC Santa Barbara: Center for Spatial Studies.* http://dx.doi.org/10.25436/E28G6D Retrieved from https://escholarship.org/uc/item/5zt0p1ft.

Biographies

Kristina Bratkova is an *Early Career Researcher* with an academic background in mathematics and music, but she recently transitioned to the world of data science and is currently working as a Data Scientist Intern at the Leeds Institute for Data Analytics. Kristina's research interests are urban analytics, GIS and everything R-related.

Eric Wanjau is an *Early Career Researcher*, currently undertaking a Data Science internship at the Leeds Institute for Data Analytics. He holds a BSc in Electrical and Electronic Engineering (2021) and his research interests span domains such as robotics, computer vision, signal processing and more recently urban analytics. Eric is keen on carrying out real-world relevant research for the public good.