Creating analysis-ready open data products for open city and transportation planning

Bruno dos Santos, Anastasia Soukhov, Antonio Paez

## Submitted to The Interesting Data and Methods Session (AAG Annual Meeting 2025)

## Abstract

Transportation and urban studies research have been in a perpetual crisis in terms of reproducibility. The state of the practice in these disciplines still is to rely on research that is unreproducible, a situation that masks anything from a lack of education in open science (in the best of cases) to academic fraud (at worst). The rise of ever bigger data and unsupervised machine learning techniques can only compound this state of affairs, which prompts a need to address this state of perpetual crisis. The solution to this is the adoption of open science practices. As a collection of actions that make scientific processes transparent and accessible, these principles can add trustworthiness and reliability to our field. Thankfully, tools and workflows that are open continue to mature, from how to prepare open data products to creating reproducible papers. Our objective in this presentation is to concentrate on one aspect of reproducible research, namely the creation of analysis-ready open data products. Over the past few years, we have landed on a workflow that has worked well for us and could be useful for others in the field. Particularly, we will share the work we have done in packaging analysis-ready open data products using the R statistical language and GitHub workflows. We use two data-packages as examples, TTS2016R and CommuteR. We demonstrate the workflow to package them, their associated data papers that allow them to be cited in their own right, and demonstrate how they can be used in analysis within other manuscripts. We conclude by outlining what we consider best practices when creating open data products.