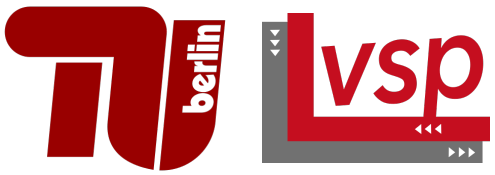

DATA SCIENCE
Verkehrssystemplanung und Verkehrstelematik
Institut für Land- und Seeverkehr (ILS)
Technische Universität Berlin

Mid-semester assignment

The goal is exploratory data analysis. See what interesting things you can find in the data set, and plot the data in a way that conveys or reveals those findings!

- Five slides/images per group (preferably groups of 2; in case you cannot find a partner: Please talk to us on 25.11)
- Please provide your slides as PDF.
- We will stitch everyone's work together into one quick-moving presentation. You will have just seven minutes to describe your findings!
- After each presentation, there will be a few minutes during which we will ask questions about your presentation.
- Remember: As this is a **pass/fail** class, your presentation will not be graded. But in order to obtain a **pass** at the end of the semester, you must present.

Timeline

- **SUNDAY - November 24th, 2024** Deadline to find a partner (Preferable group size : 2)
- **FRIDAY - December 13, 2024** Slides and code due via email to Jakob (rehmann@vsp.tu-berlin.de), Sydney (paltra@vsp.tu-berlin.de), and Billy (charlton@tu-berlin.de)
- **MONDAY - December 16, 2024** Presentations (If you cannot make it on 16.12, you must let us know by 25.11).

Objective

The purpose of this assignment is to use your R data wrangling and visualization skills to analyze MATSim output data. You will present your workflow and outputs in a short presentation (16.12). Use your transportation and/or modeling experience to come up with an interesting research question. Two different approaches would be:

1. Examine the impacts of a policy by comparing a policy case to a base case (e.g., car-free Berlin). You could, for instance, analyze the impacts in terms of score, travel time, etc. and/or examine which subpopulations are affected.
2. Like all models, MATSim results depend on (a) the input data and (b) the assumptions and mechanics of the model. Critically explore how (a) and/or (b) impact the output using your transportation know-how and (potentially) external data sources. For example, you could analyze the interrelation of different sociodemographic attributes of MATSim agents as well as the relation between these attributes and travel behavior.

Data Sources

The goal of this assignment is to analyze MATSim output data. Feel free to incorporate other data as well. If you would like to primarily use a non-MATSim data source, please talk to us.

- MATSim Berlin - Base Case: `https://svn.vsp.tu-berlin.de/repos/public-svn/matsim/tutorial/datascience2024/matsim_outputs/output-1pct/base/`
- MATSim Berlin - Carfree Scenario: `https://svn.vsp.tu-berlin.de/repos/public-svn/matsim/tutorial/datascience2024/matsim_outputs/output-1pct/policy/`

- Outputs from your own MATSim Scenario (from a previous course)
- If you're interested in the output data from a particular MATSim city/scenario, feel free to ask us.

Requirements to pass

We encourage you to use all the tools necessary to address your research question. The following constitutes a bare minimum:

- Read in a data set (readr).
- Wrangle the data: You must use mutate(), group_by(), summarize(), filter(), and three more functions from the following packages: tidyr, dplyr, stringr, lubridate, forcats.
- Join multiple data sets.
- Produce at least two different types of ggplots which you will share during your presentation.
- Include some code snippets in your presentation.