

MSc ENR IUNR ZHAW
n.MA.ENR.PTED.18FS/2019.FS.001F
FS2024

Specialized MSc in GIScience
GEO 880
FS2024

Introduction semester project

Zurich University
of Applied Sciences



Patrick Laube

Geoinformatics Research Group

Institute of Natural Resource Sciences
ZHAW Zurich University of Applied Sciences
Grüntal, Postfach, CH-8820 Wädenswil

patrick.laube@zhaw.ch



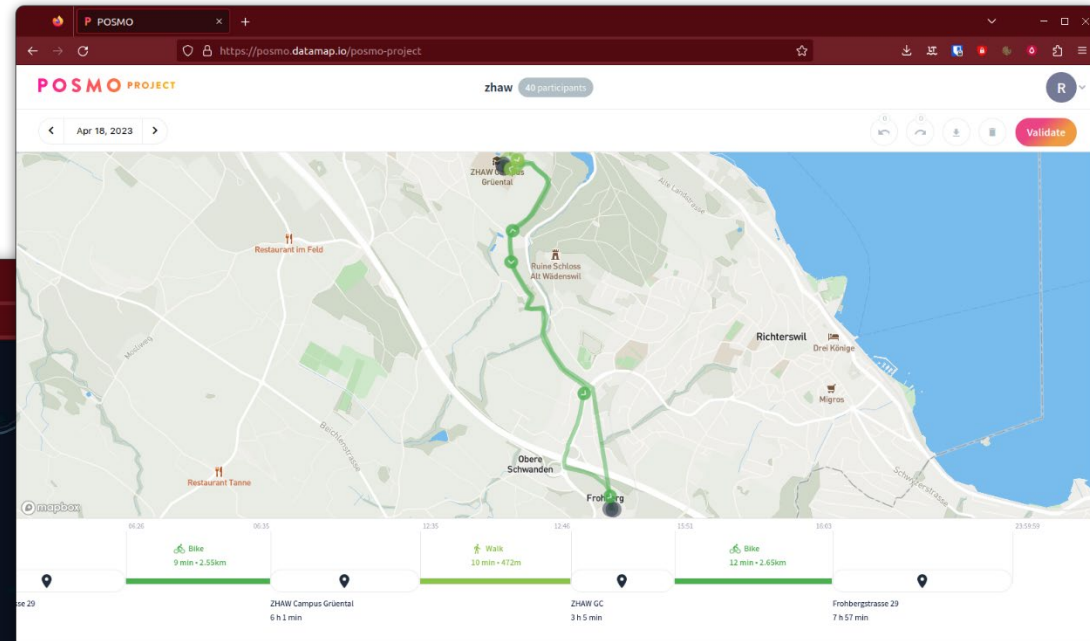
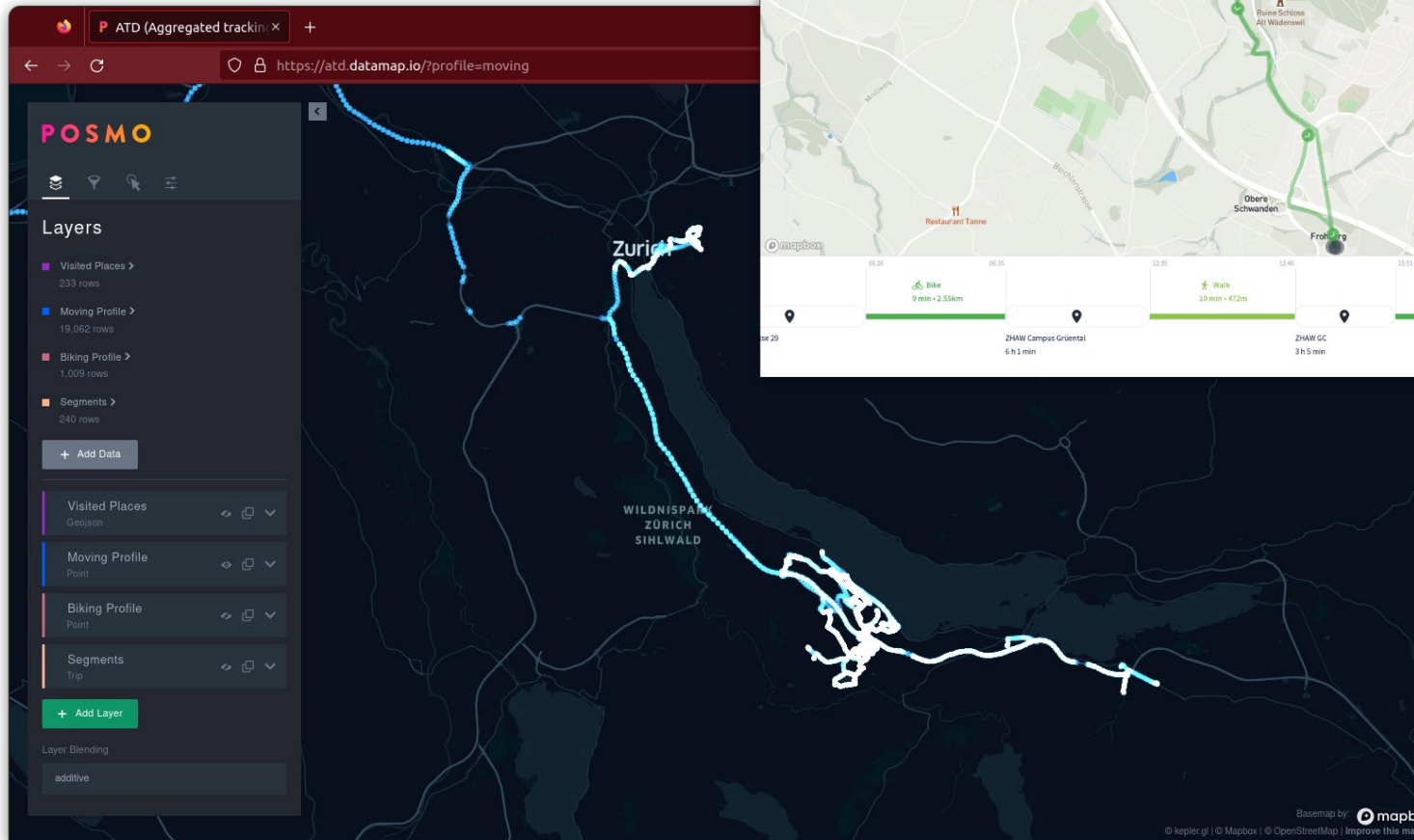
**University of
Zurich^{UZH}**

Patrick Laube

Privatdozent Geographic Information Science

Department of Geography
The University of Zürich
Winterthurerstrasse 190, 8057 Zürich

Semester-Project



Learning objectives Semester Project



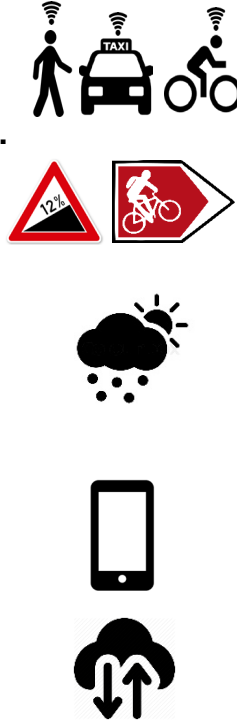
You develop data science skills that you can transfer to your own projects requiring data processing. Specifically you develop the following skills:

- ✓ You develop exploratory **data science** skills, where you iteratively develop hypotheses about your data and support your hypothesis with quantitative evidence and visualizations.
- ✓ You are able to parse, process, structure, filter, aggregate, and visualize your **data** in the software environment R and tidyverse.
- ✓ You can design and implement basic data science procedures (let's call them **algorithms**) with R, in accordance with the theory.
- ✓ You can **critically evaluate** a given data source, discuss its limitations and potential suitability for the given analytical tasks.

Semester-Project



- Travel mode detection
- Street type preferences, e.g. Slope, diurnal patterns
- Link to weather data, e.g. street preferences vs. weather, commuting patterns vs. weather
- Primarily use the data you collected yourself
- We'll set up a pool with all your (shared) data



► Show the code

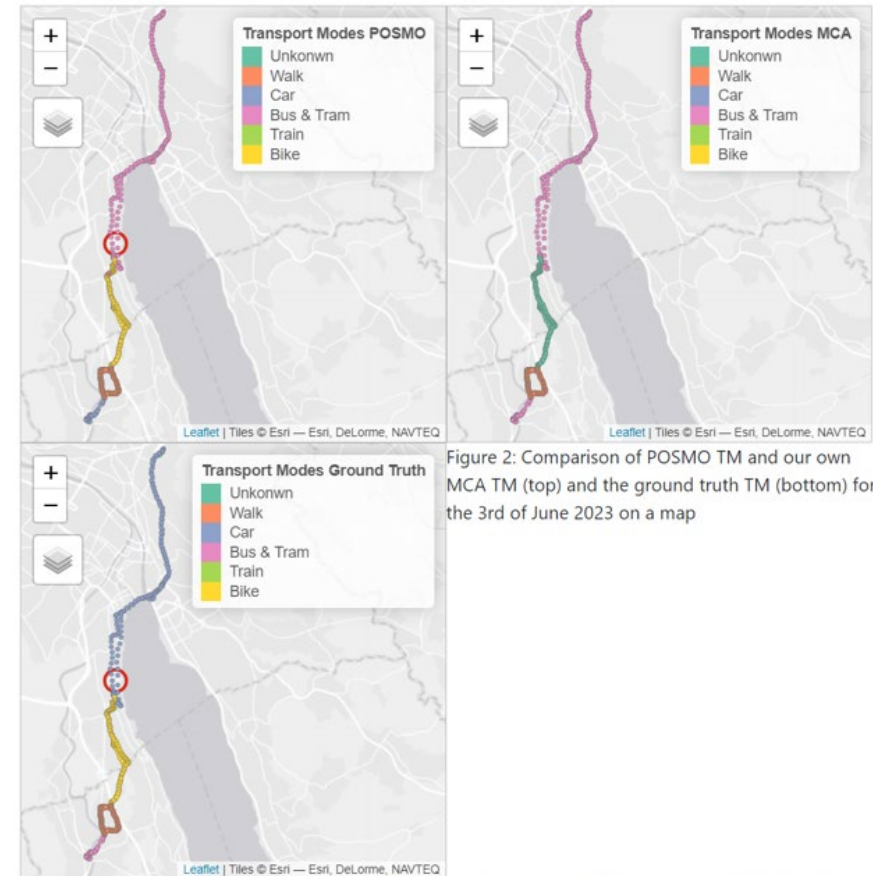


Figure 2: Comparison of POSMO TM and our own MCA TM (top) and the ground truth TM (bottom) for the 3rd of June 2023 on a map

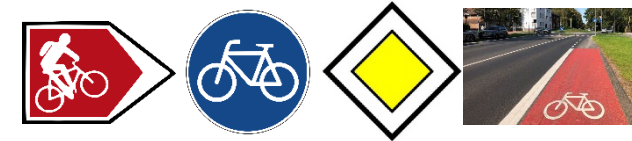
The TMD as specified in this project is overall less accurate than to the POSMO-integrated TMD when compared to the validated data, with 45% vs. 60%.

Improved Transport Mode Detection
Semesterproject, Lukas Bieri & Valentin Hett (FS23)

Possible Research Avenues

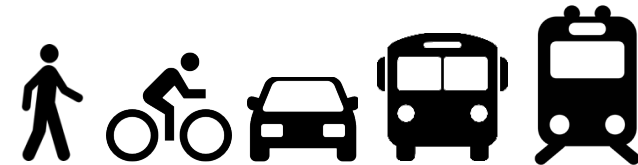
1. Street type preferences

- Diurnal patterns? Prefer certain types in morning/evening?
- Prefer certain types when climbing/descending?



2. Travel mode

- Identify travel mode from your data
- Validate against travel mode from Posmo



3. Weather impact

- Relate your mobility patterns to weather data
- Can you show an impact on the weather on the movement patterns (e.g. street type preferences, diurnal patterns, ...)



4. Difference between bike and e-bike

- Can you find different mobility patterns between bikes and e-bikes?
- Regarding slope? Distance travelled? Street types?

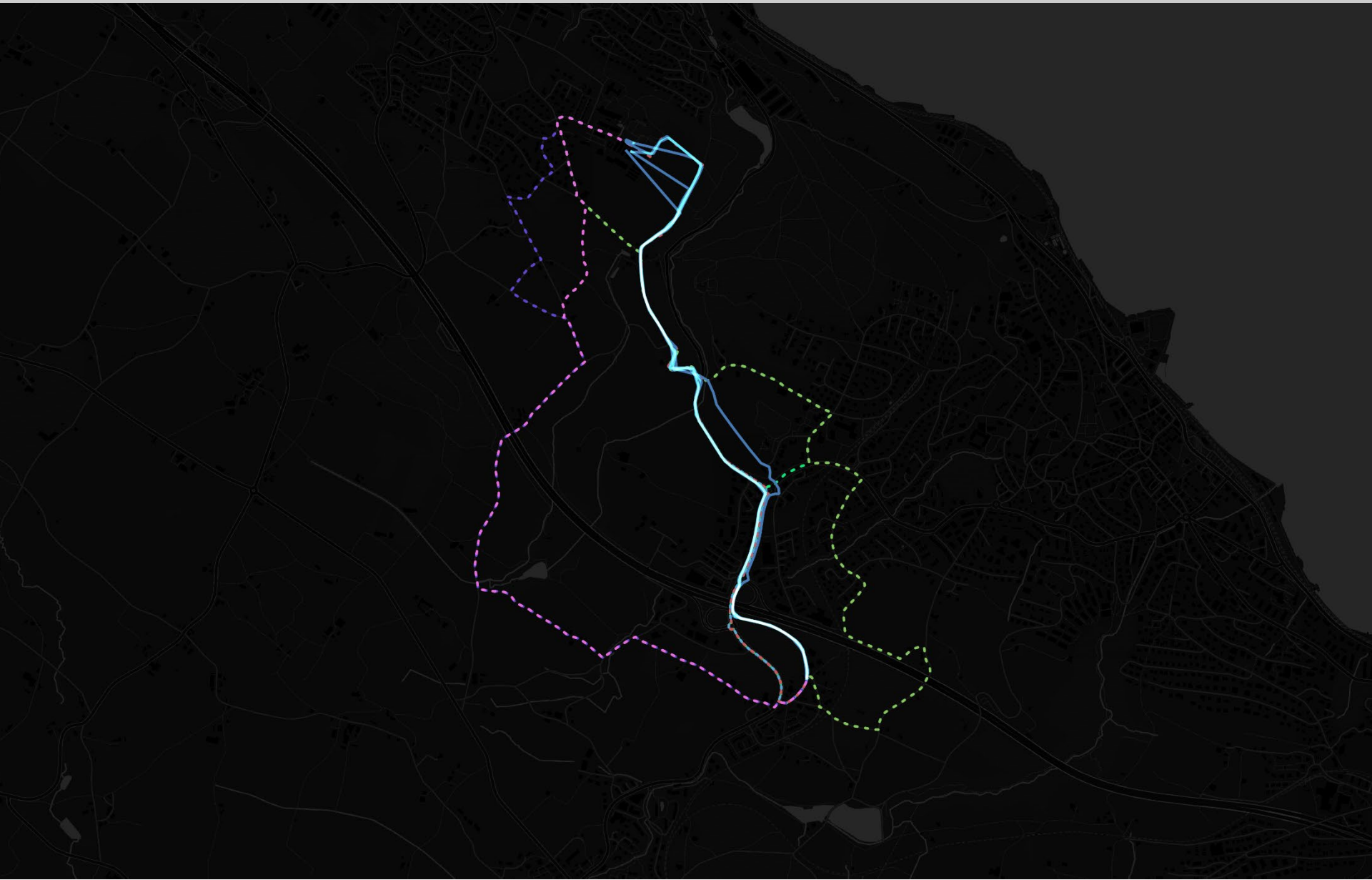


All projects

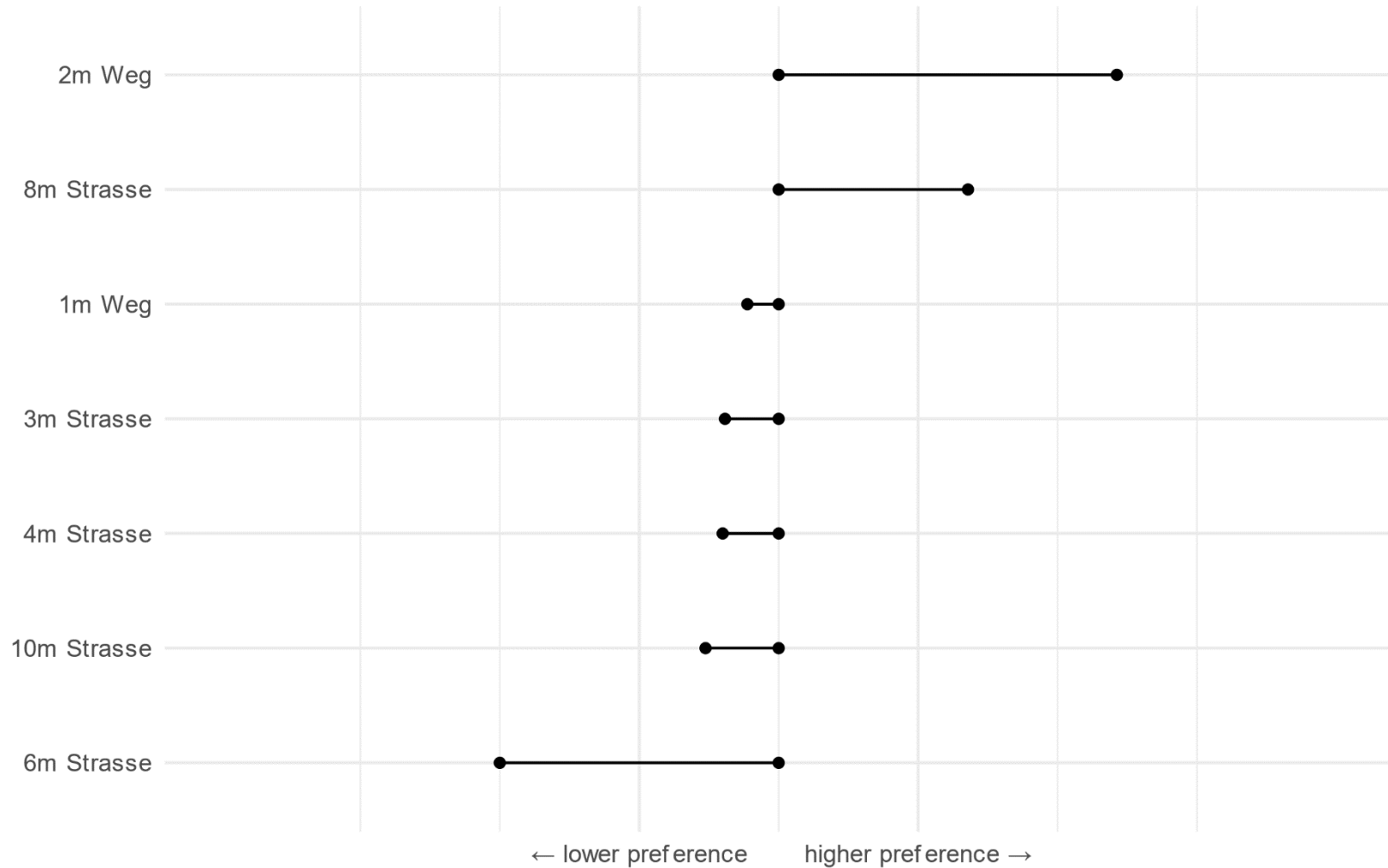
When deciding on a RQ consider the following:

- Make sure the RQ is suited to your data (e.g., do you have ascends / descends?)
- Data Preprocessing and EDA is necessary and a lot of work:
 - With trackers, separate stops from moves.
 - Do you have a mix of different transport modes that you might have to differentiate first
 - Group locations into trajectories
 - Filter out obviously erroneous trajectories/segment
- Movement data is sensitive
 - If you are willing to make your report publicly visible, make sure no private information is shared (e.g. obfuscate home location, how to do this?)
 - Even if your report and your code is publicly visible, your data does not have to be (".gitignore")

Determine preference: Possible routes (dashed) vs travelled routes



Determine preference: Possible routes (dashed) vs travelled routes



Your tasks

1. Build **teams of two** and add the team to the list on Moodle



Build Groups for Semester Project

2. Formulate at least two **research questions** for your project

- Typical types of research questions for a methodological project start with “How can this and that be conceptualized/modelled/implemented...?” or “Can this and that pattern be efficiently and effectively detected in this and that data...?”

3. Develop a **research plan** for investigating the above research questions.

- What data do you need?
- Where do you get this data from?
- What tools do you need?
- Will you use extensions of tools you got to know in E1 – E5?
- Are there other R tools or even tools beyond R that you will be using in your project?

4. Summarize your research plan in a semester **project proposal**.

- Use the template given via github
- Submit your proposal until **26.05.2024**

5. Book and attend **coaching session**

- In this coaching session you will discuss your research plan and proposal with your tutors.
- Timeslots will be made available between **27.05. – 31.05.2024** (Doodle link will follow)

Your tasks

6. **Design and implement** your data science routines that allow you investigating your research questions.
 - You may do this preferably using R or any other tool that may help you (e.g. a GIS, or FME).
 - Present your results in plots, tables, and maps.
7. Prepare a short 3–4 minute **end-of-term presentation** introducing your project to the class for **Friday 07.06.2024**. Your presentation should at least cover the following issues:
 - Context of your project (application area, conceptual models, data structures, available data)
 - Research questions and best possible outcome of your project
 - Research plan
 - Work in progress, preliminary results
 - Encountered problems and ideas for plan B
8. **Finalize your project**
 - Include feedback from the coaching sessions and the end-of-term presentations


Your tasks

9. Report your work in a written **project report**. The report has two functions:
- It shall serve you as documentation of what you did, such that at a later stage you can use the report in one of your own projects.
 - It will be used to evaluate and mark your project.

Your report shall:

- cover how you went about investigating your research questions. Describe your data science ideas and how you implemented your ideas.
- present the **results** of your study and discuss them in the light of your research questions. What have you achieved and what would be further steps for future research?
- report **problems and limitations** you encountered along the way and the solutions you chose to overcome these, be it limitations with respect to the data sources, the tools or any other source of limitation.
- discuss your data science choices in the light of the theory covered in the lectures, group works, and your reading assignments.

Project schedule

| zhaw | | | | Patterns and Trends in Environmental Data MSc in Environment and Natural Resources FS24 | | | | Geo880 Computational Movement Analysis Specialized Master in GIScience | | | |  University of Zurich ^{uzh} | | | |
|---|-------|------------|---|---|---|--|---|---|----|--------------|------------|---|---------|------------------------|-------|
| Fr 08:00 - ca. 12:00 From 19.04.2024 | | | | | | | | Patrick Laube (laup@zhaw.ch), Nils Ratnaweera (rata@zhaw.ch) & Dominic Lüönd (luoe@zhaw.ch) | | | | | | | |
| Week | Room | Date | Lesson Title | Data Science Topic | Case Study: Computational Movement Analysis | Readings Theory | Readings Skills Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly. | Reading Assignment | | Theory-Recap | Group-work | R Exercises | Project | Submission/ Assignment | |
| 16 | GA215 | 19.04.2024 | L1 Modelling matters! | Modelling means abstracting the world Implications of modelling decisions data exploration with tidyverse | Introduction into movement ecology and computational movement analysis Conceptual data models for movement spaces and movement traces | R1.1 Laube (2016). Representation, trajectories, AAG Encycl R1.2 Laube (2014). Computational Movement Analysis, Springer, Chap.1, p. 1-7. R1.3 Laube (2014). Computational Movement Analysis, Springer, Chap.2, p. 9-27. | RS1.1 Preface (16p, ix-xxiv) RS1.2 Chap2 Workflow basics (3p, 37-39) RS1.3 Chap4 Workflow scripts (3p, 77-79) RS1.4 Chap6 workflow projects (6p, 111-116) RS1.5 Chap8 Data Import with readr (21p) RS1.6 Chap13 Date and Times with lubridate (18p, 237-256) | R2.1 R2.2 RS2.1 RS2.2 RS2.3 | T1 | Gw1 | E1 | | | | |
| 17 | GA215 | 26.04.2024 | L2 Data issues | Import and preprocess data Scale, Multi-scale analysis Filtering and aggregating data Data quality: Accuracy, Precision, Uncertainty | Deriving movement parameters (speed, acceleration, sinuosity) MapMatching | R2.1 Laube, P., & Purves, R. S. (2011). How fast is a cow? cross scale analysis of movement data. Transactions in GIS, 15(3), 401-418. R2.2 Fisher et al. (2004). Where is Helvellyn? Fuzziness of multi scale landscape morphometry, Transactions of the Institute of British Geographers 29.1 (2004): 106-128. | RS2.1 Chap3 Data Transformation with dplyr (31p, 43-76) RS2.2 Chap10 Relational data with dplyr (21p, 171-193) RS2.3 Chap14 Pipes with magrittr (6p, 261-268) | R3.1 R3.2 RS3.1 RS3.2 RS3.3 | T2 | Gw2 | E2 | | | | |
| 18 | GA215 | 03.05.2024 | L3: Reproducible and Publicly Visible Data Analysis | Literate Programming, Version Control, Websites, Git, GitHub | No additional input | R3.1 Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226-1227. R3.2 Valdez (2020): Making reproducible research simple using RMarkdown and the OSF. In Social Computing and Social Media . Springer International Publishing. | RS3.1: Chap28 Quarto RS3.2: Grolemund (2021) RS3.3: Bartlett (2016): Git | | | | | | | P Intro | E2/E3 |
| 19 | | 10.05.2024 | Ascension-Friday | no class | | | | | | | | | | | |
| 20 | GA215 | 17.05.2024 | L4 Similarity and segmentation | Partitioning and structuring data Tidy data Comparing data | Stops and moves Trajectory segmentation Similarity measures and clustering | R4.1 Tao et al., (2021). A comparative analysis of trajectory similarity measures, GIScience & Remote Sensing 58.5 (2021): 643-669. R4.2 Toohey, K., & Duckham, M. (2015). Trajectory similarity measures. SIGSPATIAL Special, 7(1), 43-50. | RS4.1 Chap1 Data visualization with ggplot2 (31, 3-35) RS4.2 Chap5 Exploratory Data Analysis (28p, 81,109) | R5.1 R5.2 RS5.1 RS5.2 | T4 | Gw4 | E4 | P | | | |
| 21 | GA215 | 24.05.2024 | L5 Defining and detecting patterns | Data mining and knowledge discovery in database Algorithmic pattern detection | Movement patterns (flocks, convoys, leadership, interaction, disturbance patterns) | R5.1 Laube (2014). Computational Movement Analysis, Springer, Chap.2, p. 29-58. R5.2 Dodge, S., Weibel, R., & Lautenschütz, A. K. (2008). Towards a taxonomy of movement patterns. Information Movement Ecology, 1(1), 3. | RS5.1 Chap15 Functions (19p, 269-289) | R6.1 R6.2 | T5 | Gw5 | E5 | P | | E4 | |
| So. 26.05.2024 | | | | Submission Project Proposal | | | | | | | | | | P | Subm. |
| 27.05.-31.05 | | | | Project coaching sessions | | | | | | | | | | P | |
| 22 | GA215 | 31.05.2024 | L6 Context, Context, Context! | Semantics Spatial information Relating data | Context-Aware Movement Analysis, Semantic annotation of movement patterns, relating movement patterns to additional environmental data | R6.1 Grolemund, G. (2021). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly. R6.2 Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S. C., Kays, R., Douglas, D., ... & Wikelski, M. (2013). The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. Movement Ecology, 1(1), 3. | | R7.1 R7.2 | T6 | Gw6 | E6 | P | | E5 | |
| 23 | GA215 | 07.06.2024 | L7 Multi-Source Data | Data integration and data fusion | Combining multiple movement sensors (GPS, accelerometer, gyrometer) | R7.1 Williams, T. M., Wolfe, L., Davis, T., Kendall, T., Richter, B., Wang, Y., ... & Wilmsers, C. C. (2014). Instantaneous advantage of felid sneak -85. | | | | | | | P | Ppres | |
| So. 23.06.2024 | | | | Submission project work | | | | | | | | | | P | P |

Proj. Intro today

Proposal 26.05.2024

Coaching 27. – 31.05.2024

Presentations work in progress 07.06.2024

Submission 23.06.2024

Project Proposal

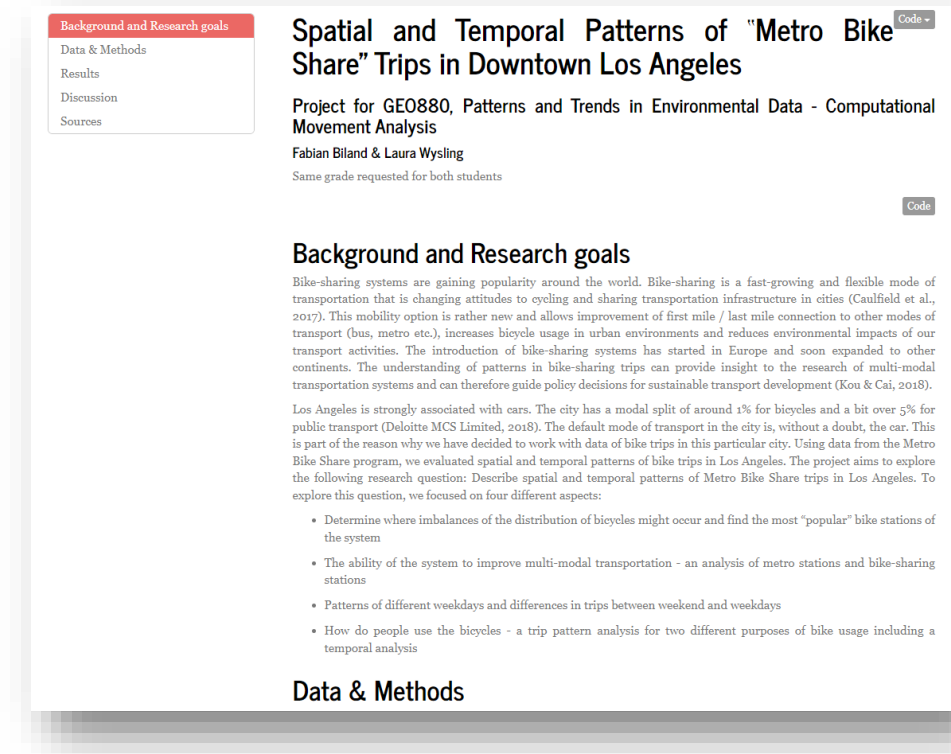
- Title
- Abstract (50-60 words)
- Research Questions (50-60 words)
- Results / products
- What data will you use?
- Analytical concepts (theory)
- R concepts / tools
- Risk analysis (with Plan B)
- What questions to us?

```
1 # Proposal for Semester Project
2
3 **Patterns & Trends in Environmental Data / Computational M
4 Analysis Geo 880**
5
6 | Semester:      | FS22
7 |-----|-----|
8 | **Data:**      | Wild Boar Movement Data
9 | **Title:**     | The title of your project
10 | **Student 1:** | Name of Student 1
11 | **Student 2:** | Name of Student 2
12
13 ## Abstract
14 <!-- (50-60 words) -->
15
16 ## Research Questions
17 <!-- (50-60 words) -->
18
19 ## Results / products
20 <!-- What do you expect, anticipate? -->
21
22 ## Data
23 <!-- What data will you use? Will you require additional co
24 -->
```

Formal requirements of the report

Your report should

- be in English or German
- be written in a **scientific writing** style.
- Length of report (approx. 15000 char (incl. spaces, incl. References list, excl. Code listing), 20000 char max)
- typically include text, maps, plots, appendices, code and references.
- be written using Quarto (see sample File in the template Repo)
- If used, use of generative AI documented & transparent?



Formal requirements of the report

Your report should

- be in English or German
- be written in a **scientific writing** style.
- Length of report (approx. 15000 char (incl. spaces, incl. References list, excl. Code listing), 20000 char max)
- typically include text, maps, plots, appendices, code and references.
- be written using Quarto (see sample File in the template Repo)
- If used, use of generative AI documented & transparent?

Marking criteria semester project Patterns & Trends / GEO 880 FS24

laup, 03.05.2024

1. Study design

- Research questions are clear, appropriate, realistic, addressable
- Are study design choices motivated from the theory CMA?
- Originality? Is it a simple extension of the exercises or are there own ideas?

2. Argument and Logic

- Students have mastered the theory and use it for their argument, with references
- Students show understanding of theory
- What has been done in terms of literature review?

3. Results

- What was achieved overall?
- How well are the results presented?
- How well are the results discussed in the light of the theory?

4. Discussion and Conclusions

- Are the conclusions drawn from the project adequate and coherent?
- Are problems that showed up been reported and alternative solutions proposed?
- Have data science choices been discussed in the light of the theory?

5. Implementation

- How was research implemented addressing the Research questions?
- What about problems and limitations and strategies overcoming these?
- Simple solutions vs. own original techniques
- Is the code properly commented/annotated?

6. Craftsmanship

- How much effort was put in documenting the work?
- Is the structure reasonable and clear?
- Length of report (approx. 15000 char (incl. spaces, incl. References list, excl. Code listing), 20000 char max)
- Is the language clear and correct?
- Are references used, correctly cited and listed?
- If used, is the use of Generative AI (e.g. ChatGPT) documented and transparent?
- Are figures and tables clear and produced to high standards?
- Overall effort and investment

Use of generative AI (e.g. ChatGPT)



- *based on Z-RL guideline AI assignments (01. 04. 2023)*
- **6.2 Use of generative AI systems in student projects. [...]** In the interests of own work or scientific integrity, the use of AI must be made as transparent as possible, i.e. it must be recognizable to third parties which parts were generated by an AI and to what extent. In principle, therefore, a declaration obligation applies to all generative AI systems that influence the quality of a work at the content level.
 - I. **A verbatim, paraphrased or analogous transfer of output** from generative AI systems must be marked at the appropriate place in a work. This applies to output without significant intellectual contribution, i.e. output that does not originate from the author's own knowledge or thoughts. The proportion or extent of the contribution of AI systems to the creative output of a work must be recognizable to third parties. The implementation follows the usual rules of scientific work, i.e. a verbatim text is to be treated in the same way as a "classic citation".
 - II. If generative AI systems are used for the **(further) processing of own data** or already generated output, the tools used must be listed in the work in summary form, e.g. in a directory, stating the purpose of use.
 - III If generative AI systems are used as **search assistants for a paper, e.g. to identify relevant literature (literature search)**, it is assumed that the learners check these references and are familiar with their content. The usual requirements for citation accuracy and thoroughness of literature references and overviews apply.


Prompt referencing 1



- <https://style.mla.org/citing-generative-ai/>

LK

In 200 words, describe the symbolism of the green light in *The Great Gatsby*.



The green light in F. Scott Fitzgerald's *The Great Gatsby* represents a multitude of symbolic meanings that are essential to the novel's central themes. The light first appears to the narrator, Nick Carraway, as a distant and alluring beacon across the water, situated at the end of the dock on the other side of the bay. The light belongs to Daisy Buchanan's mansion and is interpreted by Gatsby as a signal of his love for her.

The green light is a symbol of Gatsby's ambition, longing, and desire for the unattainable. It represents the American Dream, which Gatsby seems to elude him. The light symbolizes that they can be reunited, just as the green light symbolizes those who work hard and strive for success.

The green light is also a symbol of the corruption and emptiness that represents the shallowness and emptiness of the relationships to provide genuine satisfaction. Gatsby's pursuit of the green light and his inability to attain it underscores the status and the ultimate futility of his quest.

Overall, the green light is a powerful symbol that represents the themes and serves as a poignant reminder of the American dream.

👍

👎

Quoted in Your Prose

When asked to describe the symbolism of the green light in *The Great Gatsby*, *ChatGPT* provided a summary about optimism, the unattainability of the American dream, greed, and covetousness. However, when further prompted to cite the source on which that summary was based, it noted that it lacked “the ability to conduct research or cite sources independently” but that it could “provide a list of scholarly sources related to the symbolism of the green light in *The Great Gatsby*” (“In 200 words”).

Works-Cited-List Entry

“In 200 words, describe the symbolism of the green light in *The Great Gatsby*” follow-up prompt to list sources. *ChatGPT*, 13 Feb. version, OpenAI, 9 Mar. 2023, chat.openai.com/chat.

Prompt referencing 2



2. For a prompt e.g., "Budget for a 3 member family based in Zurich for 2023" a reference entry would be:

APA: ChatGPT. (2023, Jan. 13, 19:19 GMT). *Prompt: Budget for a 3 member family based in Zurich for 2023*. ChatGPT Jan 9 Version. <https://chat.openai.com/chat/ff75c484-3417-4149-9264-d011d456dad3>

In-text citation: (ChatGPT, 2023 "Budget for a 3 member family based in Zurich for 2023")

MLA: ChatGPT. "Prompt: Budget for a 3 member family based in Zurich for 2023." *openai.com*, 2023. ChatGPT Jan. 9 Version. <https://chat.openai.com/chat/ff75c484-3417-4149-9264-d011d456dad3>. Accessed 16 Feb. 2023.

In-text citation: (ChatGPT "Budget for a 3 member family")

3. In the case of a continuing question-answer/ conversation/ discussions/ prompts, the subsequent questions/prompts could be added after the initial question with a '/'. For example:

APA: ChatGPT. (2023, Jan. 13, 19:39 GMT). Prompt: Budget for a 3 member family based in Zurich for 2023/ Give some tips that reduce the cost in Zurich/ How to find a cost effective health insurance in Zurich?. ChatGPT Jan 9 Version. <https://chat.openai.com/chat/ff75c484-3417-4149-9264-d011d456dad3>

In-text citation: (ChatGPT, 2023 "Budget for a 3 member family ... health insurance in Zurich")

MLA: ChatGPT. Prompt: Budget for a 3 member family based in Zurich for 2023/ Give some tips that reduce the cost in Zurich/ How to find a cost effective health insurance in Zurich?. *openai.com*, 2023. ChatGPT Jan. 9 Version. <https://chat.openai.com/chat/ff75c484-3417-4149-9264-d011d456dad3>. Accessed 16 Feb. 2023.

In-text citation: (ChatGPT "Budget for a 3 member family ... health insurance in Zurich")

Hossain, Zakir. (2023). Citing and referencing ChatGPT responses: A proposal.

https://www.researchgate.net/publication/n/367091513_Citing_and_referencing_ChatGPT_responses_A_proposal

Questions?

