PA1 - Data gathering methodologies: case study, design and analysis

CIV8760E - Transport data management Frédérick Chabot & François Belisle

September 8^{th} , 2023

This first practical assignment focuses on data collection methods in the field of transportation, and aims to familiarize you with the various parameters that characterize them. It is structured around two mandates:

- 1. Case study within an existing framework;
- 2. Case study development.

1 Case study within an existing framework

For this first assignment, you will be asked to carry out certain analyses using a database based on a recent study carried out at Parc Jean-Drapeau. The study concerns the use of certain sub-spaces at Parc Jean-Drapeau. In particular, the Société du Parc Jean-Drapeau is looking at whether or not the animation and furnishings installed in these spaces encourage their use (or whether people are just passing through), and in what way(s) exactly (what activity(ies) are carried out there). The following documents (available on Moodle) will help you to better understand the context and methodology in order to produce your analyses:

- Volunteer's Guide;
- Parc Jean-Drapeau management plan presentation;
- Presentation of the methodology of studies of life in public spaces (ÉVEP) by the Centre d'écologie urbaine de Montréal;
- Presentation on the use of the mobile application for observing life in public spaces (OVEP).

This last presentation will not only help you to understand some of the limitations of the analyses you can make with the results collected by this tool, but also to be able to use the tool for the second part of this practical work.

Please note that all data manipulation for this part of the assignment must be done using SQL (mandatory). Each query used to extract the necessary information must be specified in the report. Please refer to the example below for the expected format of queries:

SELECT * FROM thisTable GROUP BY oneColumn

You can't use simple screenshots of your queries. Mapping must be done using QSIS software (or equivalent). Finally, you are free to use any tool you like to create the charts/tables (Excel, Python, etc.).

1.1 Data analysis

For the analysis, you must provide a graphical representation of a temporal analysis, a spatial analysis and a multivariate analysis (more than 2 variables). You must provide a minimum of 2 graphs for each type of analysis (6 in total). Comment on each one and make links with the methodology and/or the course material.

In addition, you need to carry out an analysis of the "Group type" variable. In other words, you need to propose a method for transforming this categorical variable into a numerical variable. Using the "Exact number" variable, compare the results obtained with your method. Justify any perceived differences.

1.2 Data model

After reading the study, please indicate whether the tables (sheets) in this spreadsheet for data entry comply with the three standards. Indicate whether or not the database follows the three standards. Finally, please present schematically what you think the appropriate model would be, taking care to note the various associations, dimensions, cardinalities and functionalities. In addition, summarize in a table (or several tables) to which entity each database variable corresponds. each database variable corresponds to.

2 Conception of a case study

As you've seen in class, there are several tools available for collecting data. Choosing the right tool is therefore an important step in data collection. This choice may depend on the limits in place (budget, time frame, etc.) and the characteristics to be collected. For this assignment, you'll need to design a case study in which you'll need to collect data by direct observation. The type of study follows that of the first part of this practical work.

2.1 Site, problem, hypothesis(es) and objective

You must then choose and describe a site, a problem (invented) and an objective that makes the case for your collection.

For example, you might decide to take Montreal's Place Shamrock, located near the Jean-Talon Market, and check out the use of public benches/tables, as in your personal experience, there seems to be a lack of seating at certain times of the day. It would therefore be interesting to know if this is the case, at what time, if one age group or gender is more affected, etc..

2.2 Tool selection

Depending on your preferences and what you have access to, you must first choose between the paper-and-pencil method and the mobile application (see the various documents posted on Moodle in the TP1 section).

2.3 Variable/attribute selection

You must choose at least 3 characteristics/variables to observe, plus a 4th, which is the geolocation of the observation made (*Mapping*). Note that for the geolocation, you can observe by "subspace" in your space to make a collection taking into account the spatial aspect. For example, think of a chessboard as a representation of your space, and an observed person might therefore have certain attributes (gender, age, etc.), in addition to doing a certain activity in the D5 zone (square) of this chessboard.

2.4 Presentation of methodology

You'll need to present your grid(s) in table, image or diagram form (depending on the tool you're using). In addition, you need to make a choice about the number of grids, the period of time to observe each grid, whether to repeat the collection over several days or weeks, etc. Please justify all your choices in terms of the parameters mentioned above (variables, site, problem, objectives, etc.). It is important in this mandate to have collected for at least 30 minutes in total in a space with more than 2 people per minute (60 people observed minimum).

This step will enable you to analyze your results (make links and expose limits).

2.5 Summary analysis

If you have carried out any data processing/cleaning, please talk about it and give clear justifications.

You should present the results in the form of aggregated descriptive statistics, and include figures (bivariate and multivariate analyses) and mapping to help understand the results. Your results should provide an understanding of the type of space and its use.

In view of the results obtained, analyze whether the parameters of your collection were appropriate to the context of the study (traffic in the space, chosen collection format, characteristics analyzed, expectations according to the space, etc.). Be transparent about any problems encountered during and after your collection, and propose solutions (other tool, size of space, etc.). Describe the advantages and disadvantages of your solutions compared to what was done here.

3 Modalities

For the second mandate, you must have collected at least 2/3 of your observations before the second laboratory (September 22) to begin your analyses with the observations you will have.

This work is done in groups of two or three. You can access this file to enter your team number and personal information. A **report** in PDF or Word format, not exceeding **30 pages**, must report on the mandates of this practical assignment. The due date is 11:59 p.m. on October 5, 2023. The file must be submitted in electronic format on Moodle.

The name of the file must include the following nomenclature: EQ{team number}_TP{number of TP}_{semester of study (A, H or E)}{year of study}. For example, "EQ01 TP1 A23".

Particular attention will be paid to the writing (French mistakes will be penalized, as will poor general organization of the work), counting for 5% of the assignment.

Please consult the Guide de rédaction pour ingénieur civil available on Moodle in the Resources section to ensure that your report is in due form.