# Introduction

Cities grow, shrink and even disappear. Nevertheless, all inevitably change, reflecting the ever-changing human society. We, people, have built our cities to accommodate our needs of shelter and social interaction. However, those needs were different 50 years ago than they are now and disparate 500 or 5000 years ago. The historical, geographical and societal differences in our needs are imprinted in every small village, town, and metropolis, leaving distinct patterns of development behind. These changes in the way how we design our cities and how their urban form is materialised can be tracked, studied, and can later influence the environment we create for ourselves today.

This thesis aims to contribute to the knowledge of urban morphology, the study of human habitat [@moudon1997], by proposing a data-driven method of analysis and classification of urban form able to distinguish the physical imprints of our needs as patterns in the built environment.

## Context of the study

There are two different perspectives when it comes to the study of patterns of development of cities. One tries to capture them to understand their influence on other aspects of life. The other tries to recognise their inner logic and processes of their formation and transformation. Neither of them is new, and both are deeply interconnected.

The study of relation to patterns of urban form is present in a wide range of fields. Economists are interested in the effects of density [@ahlfeldt2019], a social scientist may look into social mobility [@ewing2016] sustainability [@bramley2009urban] or include aspects of urban morphology into geodemographic classification models [@alexiou2016]. The role of form is present in energy consumption research [@ewing2008impact; @banister1997sustainable] or study of biodiversity [@tratalos2007urban; @andersson2014understanding]. The list could go on. What all have in common is an attempt to understand the consequence of planning decisions and hence influence the future shape of cities. What this perspective needs is a complex characterisation of urban form which does not limit it to one or few particular aspects easy to capture. For that, it needs tools and methods which are universal enough and easy to use and interpret [@boeing2020]. Urban morphology, an interdisciplinary study of urban form, focus on such characterisation. However, its tools and methods are not always optimal for the changing needs of today’s research.

Urban morphology as a specific field of research was formally established in the early 1960s in the work of MRG Conzen [@conzen1960], a geographer, and independently in the work of Saverio Muratori [@muratori1959studi], an architect. The stretch between geography and architecture is typical for urban morphology and forms the core of its interdisciplinarity. Since then, the discipline expanded and proposed different approaches [@oliveira2016; @kropf2017], some positioned far from the original qualitative works focused on processes and longitudinal aspects [@porta2006; @hillier1996; @batty1997; @batty1987]. The delineation of patterns of urban form has been studied from various angles, ranging from land use [@caniggia2001] to the historical origin and geographical location [@conzen2004thinking], societal form [@thienel2013stadtewachstum], building regulations [@forster1972court] and architectural layout [@beresford1971back]. In recent years, the attempts are more often including computational geography, data science and purely quantitative description of the form [@araldi2019; @berghauserpont2019; @dibble2017; @feliciotti2017; @li2020; @mottelson2020fine; @taubenbock2020; @usui2019]. Such approach, if turned into the systematic and comprehensive method, could react to the needs outside the niche of urban morphology as resulting characterisation could have potential to be adopted by other fields seeking to understand the relation of form and other facets of life.

The rise of data-driven approaches is not coincidental. Current era offers more abundant geographic data than any other before [@singleton2019]. Satellite imagery can now bring detailed data on the change of cities at almost real-time in a resolution of 50cm per pixel or less [@planet2020; @maxar2020]. Governments and municipalities are increasingly releasing their mapping products under open licenses [@un-habitat2020] and OpenStreetMap, the largest crowdsourced mapping project is enhancing its coverage and quality, making it a reliable source of data for morphological analysis [@barron2014comprehensive; @sehra2020extending]. Data science tools to handle large geospatial datasets [@matthew\_rocklin-proc-scipy-2015; @yu2015geospark; @hughes2015geomesa] are readily available, together with general-purpose algorithms helping to make sense of the abundance of data [@pedregosa2011scikit; @abadi2016tensorflow; @paszke2019pytorch]. The age of *Big Data* might enable to build better geographical models over space and time [@gonzalez2013big], but similarly to geography itself [@singleton2019], urban morphology needs to bring new methodological tools to increase its relevancy in digital times.

The combination of data abundance, new tools and urban morphology has a potential to deliver detailed analysis on an unprecedented extent as scalable algorithms with a potential to handle *big data* can, in theory, analyse metropolitan and larger areas while keeping information on the granular level. This idea is in the heart of this thesis, and the research presented on the following pages aims to propose steps towards this goal.

## Problem statement

Quantitative (big) data-driven methods are new in urban morphology and far from being matured. While network-based approaches like Space Syntax [@hillier1996] or Multiple Centrality Assessment [@porta2006; @porta2010] have been around for more than ten years, including tools and wealth of publications, their scope is limited. Recent additions, building on the previous theories, as Multiple Fabric Assessment [@araldi2019] or street, plot and building types by @berghauserpont2019a are trying to change the situation and expand the existing scope, but there is a long way towards comprehensiveness able to capture the complexity of urban form.

In particular, a focus on delineation of homogenous patterns of development is scarce. Published literature offers a small number of methods which are data-driven and able to work on a large scale. If there are such methods, they are limited either in terms of classification detail (e.g. @taubenbock2020) or granularity (e.g. @jochem2020). Furthermore, although based on a large sample of features, methods are often based on a small number of variables, limiting their ability to deal with (variable) selection bias and complex nature of urban patterns. Methods which would delineate homogenous areas, systematically classify them and determine the relationship between different types are rare and lack some of the other aspects mentioned above.

Classification of urban form patterns into meaningful, data-driven types is in its infancy. Literature either classify features into predefined types [@lehner2019], determine relations between cases [@dibble2017; @serra2018a] or identify areas without a further interaction [@araldi2019] between them. The critical aspect which should be studied but it is not to date are relations between automatically recognised urban patterns capturing their similarity, dissimilarity and potentially even phylogenic affinity.

## Aim and scope

This thesis aims to propose a method of derivation of numerical (data-driven) taxonomy of urban form patterns. Numerical taxonomy is a specific type of hierarchical classification which is based on quantitative characterisation of samples, reflecting the relationship between them (see chapter 3 for details). Furthermore, the proposed method should be able to delineate homogenous urban patterns used as samples in the classification in an unsupervised manner, without prior specification of types to minimise the potential bias built in the definition of the types. Optimally, the resulting method will overcome some of the limitations of previous research and provide a comprehensive description of urban form, which will inclusively cover the whole urban fabric instead of predefined case samples. The inclusive taxonomy can be, from a certain perspective, seen as an ultimate classification, one which is able to allocate any urban pattern into a hierarchical structure. Learning from other scientific fields, especially biology, taxonomies are generally accepted as the optimal model of systematisation as they capture not only individual species but also their similarity and relationships. Urban morphology cannot offer such a classification at the moment and existing methods do not aim for delivering one (see the chapter 3 for details).

The scope of work is limited to a quantitative approach enabling large scale analysis and minimal data input, expanding the applicability of results. This research is purely form-focused to further reduce data requirements, excluding land-use data, points of interest (POI), or any other additional data layers generally used in urban analytics. When describing the built environment, the current abundance of data of various kinds gives us endless opportunities to use the various dataset in the analysis. However, there are two reasons to limit the inputs to the fundamental minimum at this stage of research: 1) as the title suggests, this thesis aims to develop a methodological foundation - a framework on which further research can build. Such a framework should consist only of necessary parts and allow the flexible addition of other components based on the specific needs of future research applications. That said, the inclusion of other data inputs reflecting open spaces, green and blue space or POIs should be considered in later stages and is out of the scope of this work; 2) the framework itself should be applicable across different contexts, esp. regarding varying data availability. If we base the method on a rare dataset representing, for example, the placement and size of trees in streets, the resulting method will be applicable in a handful of cities around the world that can provide such input. However, the goal is to develop a basis that can be applied in the data-rich European context and cities of the Global South with limited cartographic representation.

From the perspective of data sources, it is limited to vector representation of urban form as raster-based earth observation does not yet offer a detailed understanding of morphological elements within remotely sensed data. Nevertheless, it is assumed that earth observation will play a crucial role in morphological research in future as it rapidly evolves and by definition allows the consistent quality of data in any context on the Earth.

The aim and scope drive the background analysis presented in chapters 2, 3 and 4. Explicit research questions and hypothesis (see Chapter 5) are then formulated based on conclusion derived from each of the background chapters.

## Significance of the study

Due to its scope, this research should provide broad applicability, allowing classification of a large number of urban environments while providing their descriptive numerical characterisation. Urban areas around the world are currently covered either by governmental, crowdsources or even private data capturing elements of urban form. As the method results in a classification based on numerical profiling of each recognised urban pattern, it could become an input for the studies analysing the effect of urban form on other aspects of life in cities. This research’s role is methodological, aiming to provide a descriptive layer on which other studies can build. For example, to study the effect of urban form on obesity, a researcher typically has to characterise both aspects - urban form and obesity. However, having expertise in both is rare, often leading to the superficial description of one. This research should deliver a thorough evidence-based portrayal of the urban environment, which could be directly embedded in such a study. For that reason, this work is, from a technical perspective, designed as replicable and reproducible research, enabling an easy application by other researchers to other areas.

On a theoretical level, this thesis proposes a comprehensive morphometric description of urban form, including specification of fundamental elements and a framework for their analysis. Furthermore, it revisits the implementation of originally biological concepts of morphometrics and numerical taxonomy in urban morphology, along with the specification of classification units.

On a practical level, it provides a methodological foundation for the construction of an expandable hierarchical classification of urban form. Moreover, it comes with bespoke software tools for quantitative analysis of urban form backing the whole research.

## Overview of the study

The thesis is structured in two major parts, *background* and *core* each with three individual chapters. The structure is graphically represented in a figure .

Chapter 2 (Existing approaches to classification of the urban form), looks into the theory of classification, introduces different methods of its application followed by the theoretical proposal of criteria for an optimal classification model of urban form. Furthermore, it provides an overview of approaches to classification of the urban form known in the literature to date and assesses each of them against the aforementioned criteria, allowing the identification of a gap within the existing research.

Chapter 3 (Numerical taxonomy) zooms into detail of one specific classification method - numerical taxonomy. The concept initially developed in biology has been recently introduced into urban morphology [@dibble2017]. However, as chapter 3 shows, aspects of the existing proposal need to be reevaluated before applying it further. Principles of numerical taxonomy are hence introduced and theoretically transposed onto the urban form, specifying the key question which needs to be resolved.

Chapter 4 (Urban morphometrics and its terminological inconsistency) dives into the realm of quantitative characterisation of urban form, providing an overview of nearly all potential measurements literature used to date. That requires to deal with the terminological inconsistency and missing framework for nomenclature and categorisation of measurable characters. Therefore chapter 4 proposes a classification and naming schema and applies it across a wide range of characters, enabling the identification of predominant ways of measuring and potential gaps.

The three background chapters are then synthesised in chapter 5 (Research design statement), which builds hypothesis and research questions on the relevant findings and proposes a framework for reproducibility for the rest of the work. Further, it outlines the selected case studies and links background and core chapters.

Chapter 6 (Morphometric elements of the urban form), the first of the core chapters, provides the basis for morphometric assessment by proposing its fundamental elements, ways of their aggregation and a coherent relational framework binding altogether. The chapter builds on fundamental elements of the urban form known from literature but proposes an implementation of morphological tessellation as a basic spatial unit instead of traditionally used plots. The ability of tessellation to reflect similar phenomena as the plot is then empirically tested together with the various models of location-based aggregation aiming to capture contextual information.

Chapter 7 (Identification of tissue types through urban morphometrics) uses the foundation proposed in chapter 6 and the database of measurable characters from chapter 4 to develop a comprehensive morphometric characterisation of urban form on the level of an individual building. The resulting hyper-dimensional description is used as an input of cluster analysis able to identify distinct types of urban tissues within an unrestricted urban fabric on a case study of Prague, Czechia.

Chapter 8 (Taxonomic relationships of urban tissues) proposes the final methodological step resulting in a hierarchical classification (i.e. taxonomy) of types of urban tissues delineated in the previous chapter. The whole method is then validated using additional variables reflecting historical origin, land use and qualitative classification of urban form. Furthermore, the transferability of the whole method is tested on another case study (Amsterdam, Netherlands) and both cases are then combined to assess the potential of extensibility of proposed numerical taxonomy of urban form.

Final chapter 9 (Synthesis) synthesises the research, discuss its character, potential application, limits and directions of further research.

Furthermore, this thesis contains five appendices, four containing supplementary information to relevant chapters (4, 6, 7, and 8) and one containing Jupyter notebooks allowing reproduction of the major parts of the work. On top of that, two additional annexes contain evidence of dissemination of the work. Annexe 1 consists of the open-source Python package momepy, which accompanies the work presented in this thesis, allows its reproducibility and lays a foundation for further morphometric research.

Notice that the structure does not contain an independent chapter dedicated to methods. That is due to the specific design of the core chapters. Each of them has its methodology dependent on the results from the previous chapter.

Structure of the thesis and allocation of chapters into parts.