DO TREES IMPROVE THE SUBJECTIVE QUALITY OF THE PHYSICAL ENVIRONMENT?

A quantitative study from two perspectives on trees: does distance to a park with high trees or many trees matter for the subjective quality of the physical environment?

Ann Lankhorst | s2524244
Supervisor: prof. dr. D. Ballas

Pre-Master Socio-Spatial Planning
Faculty of Spatial Sciences | University of Groningen

Final version | June 10, 2019

a.lankhorst@student.rug.nl

Word count: 7,597

Summary

An increasing focus on green spaces as means to improve liveability is visible in Dutch coalition agreements. It stands out that Groningen is one of the only municipalities particularly mentioning plans to plant more trees to improve the quality of the physical environment, and therewith liveability. This study investigates if, and to what extent, the proximity of trees in parks improves the physical environment. Previous studies show that green spaces have a beneficial influence on both Quality of Life and liveability. However, almost no research has been conducted on the influence of trees. The effect of tree height and tree density in parks on the "physical environment score" from the *Leefbaarometer 2.0* (Ministry of Interior and Kingdom Relations, 2016), a Dutch research and dataset on liveability, is tested through a multiple linear regression model. My research shows that there indeed is a positive and significant influence of trees in parks on the quality of the physical environment. Namely, the further away the park with trees is located from an observation point, the lower the physical environment score on this point is. However, the multiple linear regression model did not show a strong relationship because a maximum of eight percent of the variation in the dependent variable can be explained by the independent variables in this research. Therefore, no decisive conclusion can be drawn from this study.

Keywords: Quality of Life, Liveability, urban green spaces, parks, trees, tree height, tree density, multiple linear regression, physical environment.

Table of contents

1. Introduction

1.1 Background

In February 2019, the coalition agreement of the municipality of Groningen for 2019-2022 was published with the title *"Gezond, groen, gelukkig Groningen"* [Healthy, green, happy Groningen] (Gemeente Groningen, 2019). In this agreement, the life quality of Groningen's residents is the central element. This is reflected in the attention for, among others, an inclusive society, climate adaptation, and energy transition. According to the municipality of Groningen, especially green spaces and trees are indispensable for improving this life quality – also known as 'Quality of Life' –, as already indicated by the title of the agreement. Multiple studies agree with this, and demonstrate that green spaces are beneficial for mental health (Mitchell, 2013; Dzhambov et al., 2018; Mackerron & Mourato, 2013), for physical health (Sugiyama et al., 2018; De Vries et al., 2003) and life satisfaction (Bertram & Rehdanz, 2015; Ambrey & Fleming, 2013; Kaplan, 2001).

The focus on green spaces is characteristic for most of the recent coalition agreements within the Netherlands. This is presumably the result of the environmental standards of the European Union that recently became stricter (European Union, 2019). The attention for green spaces and healthy environments stand out clearly when one compares the coalition agreements of the five biggest municipalities. These five municipalities, based on populations in 2019, are (Central Statistical Office, 2019a):

- Amsterdam, 863,202;
- Rotterdam, 644,527;
- Den Haag, 537,988;
- Utrecht, 352,795;
- Eindhoven, 231,496 inhabitants.

Groningen is the sixth biggest municipality in the Netherlands with 231,354 inhabitants (Central Statistical Office, 2019a). However, the six biggest municipalities are located either in the north, the west or the south of the Netherlands. For full coverage of the Netherlands, I also looked into the coalition agreement of the biggest eastern municipality: Nijmegen, counting 176,707 inhabitants (Central Statistical Office, 2019a). Because the municipalities mentioned above are all cities, the term 'city' or 'cities' will be used from now on.

This paragraph provides a short summary of the different, and overlapping opinions that cities have regarding green spaces and trees. Amsterdam states that public spaces should be green and conducive to the health of its residents (Gemeente Amsterdam, 2018). Rotterdam (Gemeente Rotterdam, 2018) and Den Haag both emphasize the importance of trees, because *"Bomen vormen ons groene kapitaal"* [trees are our green capital] (Gemeente Den Haag, 2018, p.36). Utrecht strives for a healthy, green future (Gemeente Utrecht, 2018), which coincides with the goals of Groningen. Eindhoven wants a greener city because of air quality and the health and liveability of its residents (Gemeente Eindhoven, 2018). Finally, Nijmegen also focusses on green and wants to create more parks and green spaces for similar reasons as the aforementioned cities (Gemeente Nijmegen, 2018).

The shift towards the emphasis on green and trees in Groningen particularly becomes clear when we read the municipality's previous coalition agreement. This agreement is named *"Voor de verandering"* [For change] (Gemeente Groningen, 2014), and is primarily focussed on cooperation among residents to increase the social cohesion in Groningen. The Dutch words for *green* and *tree(s)* ('groen' and 'boom', plural: 'bomen') are barely mentioned in the entire document. It may be clear that green and trees are a key component of coalition agreements nowadays, and thus of society. This recent focus on green spaces in society is where my research is concentrated on.

1.2 Research problem

The aim of the municipality of Groningen, is to increase its residents' quality of life (QoL). It wants to do this with an eye for the character of the living environment and the attractiveness of the public space (Gemeente Groningen, 2019). To reach this goal, the city will, among other things, create more green spaces, plant more trees, and maintain the existing green and trees. However, it seems that the concept of QoL transforms into the concept of 'Liveability' by adding those environmental dimensions to the concept of QoL. After all, personal conditions and spatial conditions are the main difference between QoL and liveability (Ministry of the Interior and Kingdom Relations, 2016).

Recent Dutch research, the *Leefbaarometer 2.0* [liveability barometer 2.0], establishes five dimensions determining liveability: houses, population, safety, amenities, and the physical environment (Ministry of the Interior and Kingdom Relations, 2016). In their turn, a variety of indicators determine these five dimensions. In chapter 3 the *Leefbaarometer 2.0* is explained extensively.

The reason that my study focuses specifically on green spaces is because the indicator 'green' is not further specified in the *Leefbaarometer 2.0* (Ministry of the Interior and Kingdom Relations, 2016). As I stated before, it is remarkable that Groningen is one of the only two cities that mention trees in particular. Environmental elements like trees must somehow contribute to the dimension 'physical environment', and thus to the liveability of a certain place. This lack of definition in the *Leefbaarometer 2.0,* and Groningen's particular focus on trees, makes me wonder to what extent trees contribute to the quality of the living environment. This brings me to the central and supportive questions of this research:

Central question:

- *To what extent is the subjective quality of the physical environment in Groningen determined by the proximity of trees in the urban green space?*

Supportive questions:

- *What is known about the influence of green spaces and trees on both QoL and liveability?*
- *How can the relationship between the quality of the physical environment on the one hand, and trees, on the other hand, be determined?*

The aim of this research is two-sided. First of all, to broaden the practical and academic knowledge regarding the relation between green spaces, trees, and the quality of the living environment. Second, to evaluate if the municipality of Groningen's goals are achievable the way the city presents them, and to stimulate policy-debates.

1.3 Thesis structure

This paper is structured as follows: First, I will take you through important theories and concepts that form the basis of my research. In this second chapter, the beneficial effect of green spaces and trees on both QoL and Liveability is discussed. In the subsequent chapter, the data collection, the operationalisation of variables, and the methodology are discussed. After that, the results of this study are shown, which lead to the conclusion and discussion. In this last chapter, the most important findings of the research will be stated and linked to the theory and the introduction. I will evaluate the overall research process and the methods used. Limitations and suggestions for follow-up research are made as well. At the end of this paper, the references and appendices – consisting of statistical output and a list of abbreviations – can be found.

2. Theoretical framework

In the previous chapter I demonstrated that the municipality of Groningen currently has plans for a greener city (Gemeente Groningen, 2019). These plans arise from the desire to improve the life quality of Groningen's residents, for instance by improving the quality of the living environment. In chapter 1 I also state that this focus on the living environment, and therefore the focus on spatial elements, results in a shift from the concept of QoL to the concept of Liveability.

Despite the fact that QoL and Liveability are two separate concepts with their own definitions and indicators, the two concepts will not be discussed individually in this theoretical framework. The reason for this is simple. The link from academic research on the effect of green spaces to the concepts of QoL or Liveability is almost never established explicitly. For the purpose of clarity, I do make this connection explicit in this chapter either. The articles in this chapter will, therefore, be discussed as if they concern the same concept, and for this, the term 'liveability' will be used.

In 2016, the World Health Organization published a report on urban green spaces and health (WHO Regional Office for Europe, 2016). This document is a collection of a large number of relevant studies and thus very useful for my research. Although the report is informative and pertinent, this theoretical framework is not only based on a WHO's report, but also on other (sometimes more recent) academic literature. These studies can be divided into different themes, which all have some direct or indirect relation with liveability. The themes are: mental health, physical health, experience of life, residential satisfaction and housing prices. Below, the themes are set out and together, they form the answer on the first supportive question: *what is known about the influence of green spaces and trees on both QoL and liveability?*

2.1 Direct connections green spaces and liveability

2.1.1 The positive effect of green spaces on mental health

Numerous studies have been conducted on the relationship between green spaces and mental health in the past few decades – according to Eurostat, the EU's institution for (explaining) statistics, health in general is one of the indicators of QoL (Eurostat, 2019a). In his research, Mitchell (2013) states that working out in a green environment is better for mental health. He states that working out in open parks and forests is especially favourable. Five years later Dzhambov et al. (2018) endorse Mitchell's result. They examined the relationship between green spaces and mental health with regard to young adults in Bulgaria. Their research shows that the effect is indirect but positive. Green spaces (and blue spaces) lead to more physical activity, and they cause less irritation because there is less noise in green spaces. Kondo et al. (2018) support these conclusions, but state that the relationship between green spaces and *physical* health must be investigated deeper. This relation is central in the next paragraph.

2.1.2 The positive effect of green spaces on physical health

The relationship between green spaces and mental health seems easier to measure than the relationship between green spaces and physical health. Nevertheless, Sugiyama et al. (2018) show that accessible and appealing public green spaces *"could reduce chronic disease risk through facilitating physical activity and alleviating stress"* (Sugiyama et al., 2018, 15). This conclusion somewhat matches with the conclusion of Mitchell (2013), discussed in the previous paragraph. Earlier research of De Vries et al. (2003) also shows a positive relationship between green spaces and self-reported health. However, the researchers do not distinguish between mental and physical health. The same researcher, Sjerp de Vries, cooperated in research on this topic with Dinand Ekkel in 2017. Despite the fact that health is not specified in this research either, the positive effect of green spaces on health is obvious (Ekkel & De Vries, 2017). Furthermore, Eurostat states that one of the QoL-indicators, the natural and living environment, *"can have direct impact on the health of individuals"* (Eurostat, 2019).

2.1.3 The positive effect of green spaces on the experience of life

Multiple academic studies demonstrate a positive effect of green spaces on the experience of life, again an indicator of QoL (Eurostat, 2019a). In their study, MacKerron & Mourato (2013) designed an application for mobile phones. Every now and then, people had to answer the question of how they were feeling. The responses were linked to the participant's GPS locations. This led to the conclusion that people are feeling happier in green environments. Moreover, green spaces prove to be a beneficial influence on life satisfaction (Bertram & Rehdanz, 2015; Ambrey & Fleming, 2013).

Paragraphs 2.1.1 until 2.1.3 addressed studies that can be directly linked to liveability. Research shows that QoL-indicators are primarily studied explicitly in relation to green spaces. A possible explanation for this is that these indicators are more concrete than the dimensions of liveability. The liveability dimensions, as explained in chapter 1, are relatively abstract because it concerns categories. Therefore, other studies on the influence of green spaces can be indirectly connected to liveability. These will be discussed in 2.2.

2.2 Indirect connections green spaces and liveability

2.2.1 Green spaces and residential satisfaction

The concept of liveability seems to be closely connected to the concept of residential satisfaction, because liveability is focussed on the relationship between the environment and people's wishes and requirements (Leidelmeijer & Van Kamp, 2003). But what is it exactly, that determines residential satisfaction?

An answer to this question can be found in the research paper of Huang and Du (2015). According to their research on residential satisfaction in social housing in China, residential satisfaction is mainly determined by the neighbourhood environment, public facilities and housing characteristics (Huang & Du, 2015). This neighbourhood environment seems to be a concept corresponding the physical environment concept, and therefore those two types of environment will be considered identical. Seeing as my research is done on a grid scale of 10 by 10 meters, and will, therefore, be able to provide information about the neighbourhood scale. The question remains what the influence is of green spaces on residential satisfaction, and therewith on liveability.

There are some academic studies published on the subjective perception of green spaces and its influence on residential satisfaction. An example of this is the article of Kaplan (2001). She states that a natural view from the window increases the resident's satisfaction with the neighbourhood and the sense of well-being. Bjerke et al. (2006) add to this that the vegetation density is an important factor. These Norwegian researchers found that people prefer a moderate density of vegetation over a complete *"open scene"* without vegetation.

2.2.2 Green spaces and rising housing prices

It seems plausible that people, when they are moving, try to move to a place that they consider liveable. A residential place where they expect to be satisfied. Multiple studies found interesting proof for this argumentation. Panduro et al. (2018) tried to elicit people's preferences for urban parks. Their research shows that for every meter reduction of the distance to the nearest park, the housing prices increase. The same findings can be found in the study of Daams et al. (2016), in which the researchers state that *attractive* natural spaces up to 7 km influence property prices. In other words, people are willing to pay a higher price for their new home if this house is located near an attractive park. In fact, housing is one of the five dimensions that determine the liveability of a certain place (Ministry of the Interior and Kingdom Relations, 2016).

2.3 The influence of trees

The beneficial influence of trees on Liveability appears only from a handful of studies. Mitchell (2013), also mentioned above, concludes his research with the statement that *"regular users of Woods/forest for physical activity were at about half the risk of poor mental health of non-users"* (2013, 132). This suggests that green spaces with some density of trees, or a forest, have a positive impact on the health of people. In 2015, Kardan et al. supported this conclusion. They found *"that people who live in neighbourhoods with a higher density of trees on their streets report significantly higher health perception and significantly less cardio-metabolic conditions"* (p.1).

Tree density not only affects health, but also housing prices. Franco & Macdonald (2018) show that the housing prices in Lisbon are influenced by the proximity of larger forests in particular. It stands out clearly that these studies emphasize the effect of forests. However, unfortunately, until now not much research has been conducted on the relationship between liveability and trees.

2.4 Conceptual model Liveability and green spaces

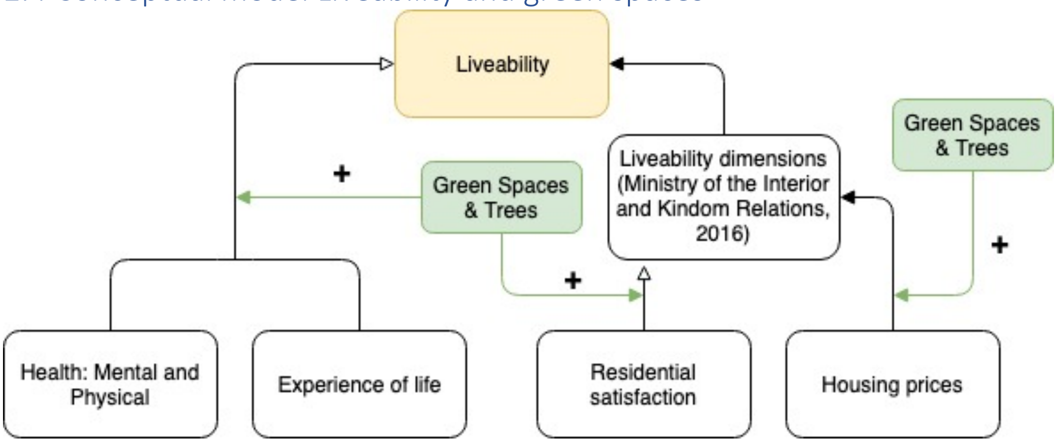


Figure 1: Conceptual model Liveability and Green Spaces. The green boxes "Green Spaces & Trees" and their green arrows, point to the beneficial influence green spaces and trees have on liveability, according to the academic literature discussed in this chapter. The black arrows represent a relationship of determination. The empty arrows represent a relationship of influence.

Despite the fact that the concepts of QoL and Liveability are not separated in this chapter, it remains important to keep in mind that there is a substantial difference between the two concepts. Especially because this has essential consequences for the data collection and analysis, which will be explained in the following chapter.

3. Data collection and methodology

3.1 Data and operationalisation

3.1.1 Data collection: case and scalar level

For this study a dataset is created out of multiple existing datasets. These datasets are managed by different Dutch (governmental) institutions, like the Central Statistical Office (CBS). Some datasets are available online for everyone, others require permission. The following datasets are merged into one dataset through GIS:

- *Leefbaarometer 2.0* [Liveability model 2.0] (Ministry of the Interior and Kingdom Relations, 2016) (see 3.1.2);
- *Bestand Bodemgebruik (BBG)* [Document Soil Use] (Central Statistical Office, 2008 & 2015);
- *Bomen in Nederland* [trees in the Netherlands] (Atlas Natuurlijk Kapitaal, 2017).

Due to ethical considerations regarding completeness and the need for control variables, some other datasets are added as well, which are:

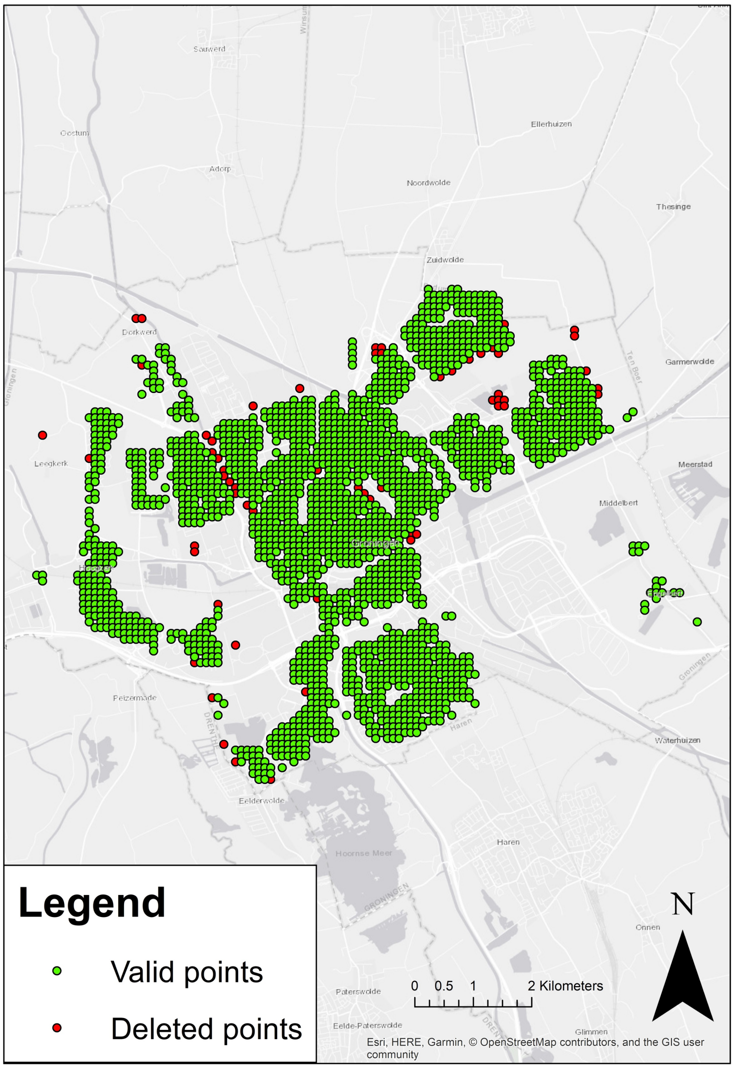
- *Kerncijfers wijken en buurten* [Key figures neighbourhoods and areas] (Central Statistical Office, 2018 & 2019b);
- Ecosystem Unit Map (Central Statistical Office, 2017a & 2017b).

The central case in this study is the city of Groningen. Therefore, only the data pertaining to Groningen is selected. The borders that are employed, are the town limits established by the dataset *Kerncijfers wijken en buurten* (Central Statistical Office, 2019b). Besides this selection, the dataset is disposed of incomplete and/or missing data.

Furthermore, the spatial scale used in the dataset is the grid scale of 10x10 meters, adopted from the *Leefbaarometer 2.0* dataset. This scale is considered most relevant and viable because the other scales on which the *Leefbaarometer 2.0* dataset is available (neighbourhoods, areas, municipalities and postal codes) can include multiple soil types or functions at the same time. For example, one neighbourhood can include living areas as well as parks (two different soil use codes, according to *BBG*, 2008 & 2015). The grid scale facilitates the separation and the clustering of different soil use areas. Therefore, a park can be considered as an urban green space. Besides that, grid scale is the lowest scalar level possible, and will consequently yield the most detailed results.

The map below, Map 1, is a representation of the observations in the research area. In this map, the observations are displayed in two kinds. The green dots are the observations included in my research. The red dots are, by contrast, the observations without available data (null values). These observations are therefore deleted from the dataset used. To some extent, the deleted dots occur in clusters along areas with less or no houses. An example is the clustering of six red dots on the upper right side of the map, where the sports centre *Kardinge* is located.

A final note must be made about the ethical considerations concerning the combining of datasets. For each dataset, it applies that the data has been gathered for specific reasons, and with specific aims. This means that the purposes of this research, might not correspond to the purposes of the studies from which the datasets are the results. This could have consequences for my research, which I should keep in mind as a researcher. However, I do not expect these consequences to be very influential.



Map 1: Valid and deleted observations in the research area

3.1.2 Dataset *Leefbaarometer 2.0*

The *Leefbaarometer 2.0* dataset can be seen as the basis of the dataset used in this research, since it provides the dependent variable. Its nature and structure will therefore be explained in this paragraph. First of all, the dataset is built upon the definition of liveability as defined by liveability-expert Kees Leidelmeijer and researcher of the Dutch National Institute for Public Health Irene van Kamp: *"Leefbaarheid is de mate waarin de omgeving aansluit bij de eisen en wensen die er door de mens aan worden gesteld"* [Liveability is the extent to which the environment meets the requirements and wishes that are set by people for it] (2003, p.59).

The *Leefbaarometer 2.0* is a biannual research on the subjective liveability in neighbourhoods and areas within the Netherlands (Ministry of the Interior and Kingdom Relations, 2016). The study is conducted by *RIGO research* and *Atlas voor gemeenten,* and resulted in the formulation of a regression model, with which the liveability can be predicted for each place in the Netherlands every two years. Because of this, there is no need any more to conduct surveys biannually.

Furthermore, the research consists of an extended dataset, which provides the dependent variable of my research (see 3.1.3). The first part of the *Leefbaarometer 2.0*-model directly predicts the resident's valuation of their living environment. The second part of the model takes housing prices into account as well for predicting this valuation. These two sub-models result in a clear picture of the liveability on certain places in the form of liveability scores ranging from -1 to +1. These scores are subjective because all predictions are made based on available information (from former questionnaires) about resident's judgement and valuation of their direct living environment.

To be able to assign a score to liveability, indexed scores are assigned to multiple indicators. These are spread over the following five dimensions: houses, population, safety, amenities, and physical environment. To each indicator a score is allocated. The sum of the scores that belong to one of the five dimensions results in a score for that dimension. The sum of the five dimension scores result in the overall liveability score.

3.1.3 The dependent variable: physical environment score

The dimension 'physical environment' from the *Leefbaarometer 2.0* is the dependent variable in my research. This dimension can vary from -1 to +1, and consists of indicators that are either beneficial to the environment, or ones that are not. According to the justification report of *Leefbaarometer 2.0*, green spaces are seen as beneficial to the physical environment (RIGO & Atlas voor gemeenten, 2014). However, green spaces are not defined or specified. It is therefore not clear if meadows are green spaces, or if there must be bushes and trees as well.

For this reason, this study investigates through a multiple linear regression model to what extent vegetation – in the form of tree height and tree density – influences the resident's judgement and valuation of their direct physical environment. My dependent variable does not include explicitly any of the independent variables or control variables used in my research. Therefore, the effect of tree height and tree density on the subjective quality of the physical environment can be statistically tested.

3.1.4 Creating new variables relating to tree height and tree density

The merging of the different datasets resulted in a map of Groningen, consisting of multiple superimposed layers. In each layer, relevant information concerning one element is included. For example, one of the layers displays all trees in the city of Groningen. In order to analyse this data, the maps are transformed from GIS into a numerical dataset. This includes, on the one hand, the physical environment scores from the *Leefbaarometer 2.0* as the dependent variable, and on the other hand several other variables as independent ones. Despite the fact that the dataset in this study consists of secondary data, the independent variables had to be made by combining three elements: soil type, tree criteria, and distance. I will elaborate on this below.

First, a choice had to be made regarding the soil types from the *BBG* (Central Statistical Office, 2008). This dataset consists of 38 categories, of which only three consist of public green spaces and might, therefore, have a positive impact on the quality of the living environment. It concerns the following types: park (no. 40), areas for day-recreation (no. 43), and forest (no. 60). Due to the time limitation of this research project, only one soil type could be included in this study.

I have chosen to include parks (*'park en plantsoen',* type no. 40) because the city of Groningen is the area of research. In my view, parks occur more often within city limits than areas for day-recreation or forests. Therefore, it seems to be most relevant to include parks and build the independent variables upon this type. With that, the 'urban green space' in the central question of this research is operationalized as parks. How the production process of variables went, is explained in the following.

After the selection of the category 'park', five interaction variables are created based on two types: the height of trees and the density of trees. To every variable applies that an area needs to be at least 1ha, the minimum surfaces the *BBG* uses as well for parks (Central Statistical Office, 2008). The first type is the height of trees and provides two variables. From all trees in Groningen, documented by the *ANK* (2017), I selected the highest 33,3% of the trees and regarded these as 'high trees' that might affect the physical environment score. The first of the two variables determines whether an area has at least 70% of high trees on 1 hectare. The second one checks whether this percentage is at least 90% on 1ha. In other words, of all trees on a certain hectare, at least 70%/90% must belong to the group of the highest 33,3% of trees in Groningen.

The second type provides three variables and concerns the tree density in certain areas. As I have explained in the introduction, the municipality of Groningen specifically mentions trees as a measure to improve the physical living environment. Therefore, this research does not focus on vegetation in general, but on trees in particular. In response to the study of Bjerke et al. (2006), the density of trees in parks is included in this research as an independent variable. Three density limits are used to create another three variables – besides the two height variables explained above – which are clarified below.

For every density variable, the principle holds that at least 0.5ha of 1ha has to adhere to a certain percentage of trees. This principle is not based on academic literature, since there is none, but is based on logical substantiation. Furthermore, the limits on which the three variables are based, are not based on academic literature for the same reason. The first of the three variables is 50% density. This means that at least 50% of 0.5ha of 1ha must be forested, which is 25% of the total hectare. The second variable represents the lower bound of 70%. The last represents a lower limit of 90%. In practice, this results in a decrease in the number of cases when the percentage raises. In fact, there are no observations for the variable density of 90%. This is not unexpected, because a park with a tree density of 90% will probably be classified as a forest in the *BBG*.

Eventually, for all these variables the distance in meters to the grid cells from the *Leefbaarometer 2.0* is calculated. Distance (in meters) is not an uncommon measure in research on green spaces (see 2.3.1). Therefore, distance is the value the five independent variables finally have in the dataset. The last step is to double the variables because the distance to the nearest park without the tree criterions has been measured as well. Table 1 below enumerates the five interaction variables. The areas without trees are not included in this table.

	Tree height	Tree density
Park	At least 70% high trees on 1ha	At least 50% of 0.5ha is forested
	At least 90% high trees on 1ha	At least 70% of 0.5ha is forested
		At least 90% of 0.5ha is forested

Table 1: The interaction variables of this study.

3.1.5 Control variables

It is important to add control variables to the linear regression models as well (Punch, 2014). Control variables are used to check whether the dependent variable is related to the independent variables, as it could be affected by outside effects as well. In this research the control variables that are used can be linked to the four remaining dimensions of the *Leefbaarometer 2.0*: houses, population, safety, and amenities. The control variables are part of the four dimensions mentioned above or are proxies for indicators of those dimensions.

Because my dataset does not contain a variable 'average income', the variable "Woz-waarde" [property value] is selected, which can be seen as a proxy for income. There are no extra variables selected from the fifth dimension (the physical environment). The reason for this is, that they are used to build up the physical environment score. They will therefore negatively affect the specification of the linear regression model. Below, the control variables are enumerated per *Leefbaarometer 2.0*-dimension.

- Dimension: houses
  - Percentage of private houses
  - Percentage of houses owned by a housing corporation
- Dimension: population
  - Woz-waarde (property/cadastral value)
  - Population density
- Dimension: safety
  - Average number of robberies in houses, sheds, etc.
  - Average number of damages and crimes
- Dimension: amenities
  - Distance to the nearest supermarket
  - Distance to the nearest family doctor

3.1.6 The standardization of variables

After running some regression models, I have decided to standardize all variables. This is due to the delivered output of these models, which consisted of values that are difficult to interpret. Table 2 is an example of this and can be read as follows: the coefficient of a model with the physical environment score as the dependent variable, and the distance in meters to the nearest park of which 70% of the trees is a high tree, is –3.88e-17. This means that, for every meter the nearest park is further away, the physical environment score decreases with –3.88e-17 (see table 2). This value is so small, that is difficult to interpret, which is why all variables are standardized. Therefore, the values of the coefficient, the robust standard error and the confidence interval are standard deviations from now on (see chapter 4).

Linear Regression			**Number of observations**	2,204
			F (0,2202)	.
			Prob>F	.
			R-squared	0.0033

Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Distance to nearest park where 70% of the trees are high trees	–3.88e-17*	1.34e–17	–6.51e-17	–1.25e–17
Constant	–.018446*	.0055788	–.0293862	–.0075057

*Table 2: Linear regression output before variables are standardized. Dependent variable: physical environment score. Independent variable: distance to the nearest park with 70% of the trees are high trees. *P≤0.05. The academic notation "–3.88e-17" is maintained to show the extent to which the output is difficult to interpret.*

3.2 Methodology

As I have shown above, in this study many interaction variables are included. These variables have continuous values in the form of distances in meters. Through multiple linear regression, the effect of parks with high trees or many trees in close proximity on the quality of the living environment is investigated. For the statistical analyses, the statistical software StataSE (version 15.0.585) is used.

The first step is to test the relationship between the dependent variable (Y), and one independent variable (X) each time. The next step is to build up the linear regression models with control variables. First, from every dimension (see 3.1.4) only one control variable is picked and put in the model. The output of every model is compared against the single linear regression outputs. The last step is to add the second control variable from every dimension as well. Again, the output is compared against the former models.

These linear regression models will result in independent variables that have a significant effect on the dependent variable (Y), and (probably) variables that are not significantly related to the physical environment. A standard significance level of 95% will be used, equal to $p<0.05$. The significant models will be discussed in the following chapter, from which conclusions are drawn.

4. Results

4.1 Descriptive statistics

Table 3 displays the descriptive statistics of the dataset used in this research. Every time, the original variable is listed first, after which the standardized variable is listed. This provides a clear overview of the data this research is based on.

	Obs	Mean	Std. Dev.	Min	Max
A	2,204	–	–	–	–
Physical Environment score from *Leefbaarometer 2.0*	2,204	−.0331001	.1789977	−5.75	5.1
Standardized Physical Environment Score	2,204	8.33e−10	1	−31.9384	28.67691
Distance to nearest park where 70% of the trees are high trees	2,204	3.78e+14	2.66e+14	0	9.96e+14
Standardized distance to nearest park where 70% of the trees are high trees	2,204	−1.55e−10	1	−1.422186	2.325233
Distance to nearest park where 90% of the trees are high trees	2,204	3.60e+14	2.72e+14	0	9.99e+14
Standardized distance to nearest park where 90% of the trees are high trees	2,204	−9.04e−10	1	−1.324114	2.351012
Distance to nearest park where 50% of 0.5ha is forested	2,204	3.20e+14	2.48e+14	0	9.97e+14
Standardized distance to nearest park where 50% of 0.5ha is forested	2,204	2.62e−10	1	−1.28597	2.72586
Distance to nearest park where 70% of 0.5ha is forested	2,204	3.83e+14	3.05e+14	0	9.99e+14
Standardized distance to nearest park where 50% of 0.5ha is forested	2,204	2.42e−10	1	−1.255625	2.024601
Population density	2,204	6793.487	4009.177	0	17548
Standardized population density	2,204	2.32e−10	1	−1.663555	2.682474
Property value/ 'Woz'-value: proxy for income	2,204	173.0676	79.3692	0	503
Standardized property value: proxy for income	2,204	−3.82e−09	1	−.7568126	4.156932
Percentage private houses	2,204	46.93013	26.12647	8	98
Standardized percentage private houses	2,204	2.49e−09	1	−1.490065	1.954718
Percentage houses owned by a housing corporation	2,204	32.50862	22.24247	0	89
Standardized percentage houses owned by a housing corporation	2,204	−1.80e−09	1	−1.461556	2.539798
Number of thefts from houses and sheds	2,204	5.549002	2.650098	0	20
Standardized number of thefts from houses and sheds	2,204	−4.25e−09	1	−2.093886	5.453006
Number of acts of vandalism	2,204	6.602541	4.421599	0	36
Standardized number of acts of vandalism	2,204	−1.16e−08	1	−1.493248	6.648604
Distance to the nearest family doctor	2,204	.8330762	.5792441	.2	4.2
Standardized distance to the nearest family doctor	2,204	−3.22e−09	1	−1.092935	5.812616
Distance to the nearest supermarket	2,204	.6498639	.5157923	.2	4.1
Standardized to the nearest supermarket	2,204	5.99e−09	1	−.8721802	6.689002

Table 3: Descriptive statistics of the dataset used in this research. Original variables are listed, as well as the standardized variables.

4.2 Model 1: Tree height

This section presents the results of the first linear regression model. As I have explained in chapter 3, five independent variables are created. Two of those concern the tree height in parks: distance to the nearest park where 70% **or** 90% of the trees are high trees. Only one of those is significantly related to the physical environment score. Table 4 shows the results of the model with this significant independent variable.

Linear Regression			Number of observations	2,204
			F (9,2194)	288.45
			Prob>F	0.0000
			R-squared	0.0861

Physical Environment Score	Coef.	Robust Std. Error	[95% Conf. Interval]	
Standardized distance to nearest park where 70% of the trees are high trees	−.040914*	.019802	−.079747	−.002080
Standardized population density	−.132120*	.012214	−.156073	−.108167
Standardized property value	.055592*	.028022	.000639	.110545
Standardized percentage private houses	.226306*	.030534	.166427	.286186
Standardized percentage houses owned by a housing corporation	.089329*	.016461	.057048	.121610
Standardized number of thefts from houses and sheds	.026646*	.011835	.003436	.049856
Standardized number of acts of vandalism	.015101	.010884	−.006242	.036445
Standardized distance to the nearest family doctor	−.108236	.068639	−.242841	.026368
Standardized distance to the nearest supermarket	.068476	.053061	−.035578	.172531
Constant	.000000	.020405	−.040015	.040015

*Table 4: Model 1. A multiple linear regression model with the dependent variable (Y), the physical environment score, and independent variable X1: distance to a park, of which 70% of the trees are high trees. *P≤0.05.*

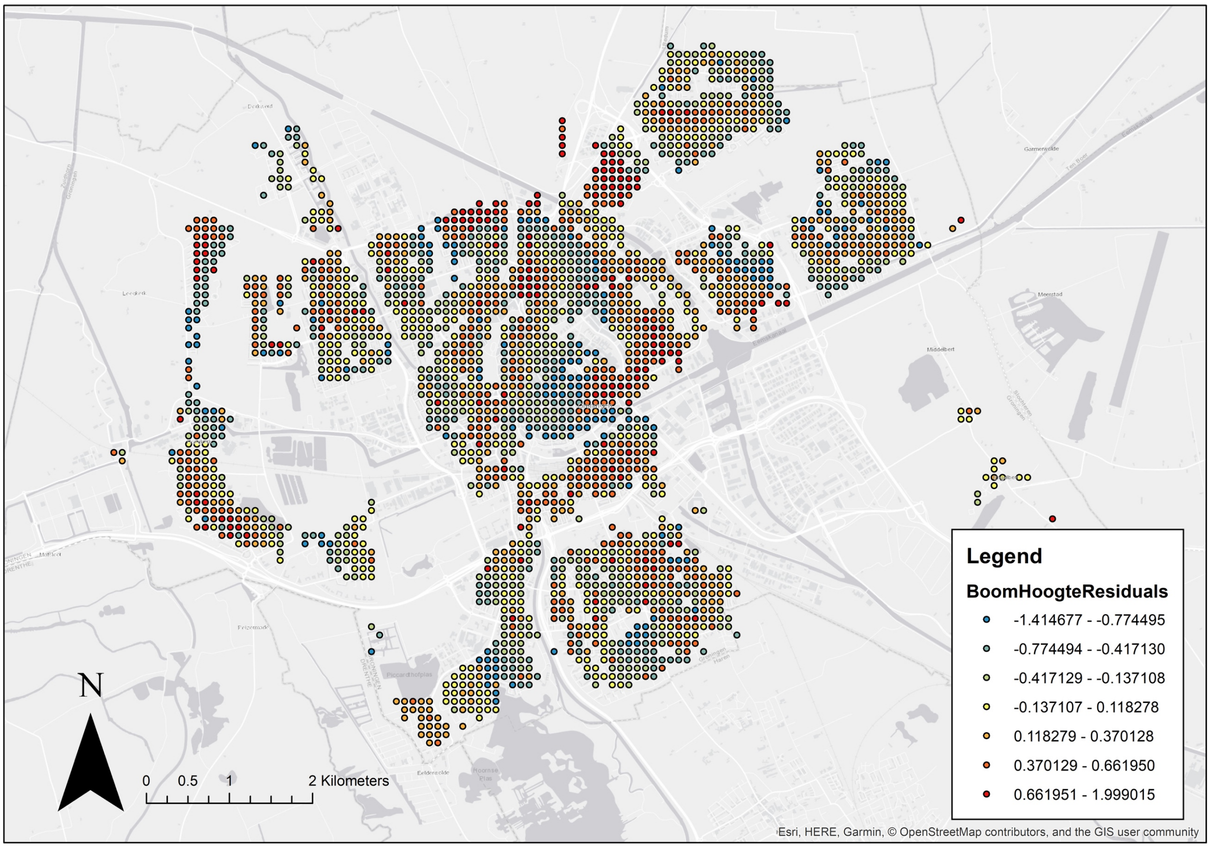
As can be seen, the independent variable “distance to the nearest park where 70% of the trees are high trees” significantly affects the dependent variable “physical environment score” (p=0.039≤0.05). Besides that, the linear regression model itself is significant as well, as is shown by the F-test result of 0.000. These values are quite similar in the linear regression models with less or no control variables (see appendices).

It can be argued that the negative coefficient of the variable “distance to the nearest park where 70% of the trees are high trees” is a positive result in the context of Groningen’s coalition agreement. The negative coefficient indicates that for every standard deviation the park with 70% high trees is located further away, the physical environment score decreases with 0.040 standard deviation. In other words, the further away the park with high trees is, the lower the physical environment score. This corresponds to the principles of the *Leefbaarometer 2.0* because the researchers of the *Leefbaarometer 2.0* consider green spaces to be ‘satisfiers’ for the physical environment (Ministry of the Interior and Kingdom Relations, 2016).

However, only eight percent of the variation in the dependent variables can be explained by the independent variable X1. This can be told by the R-squared of 0.086. This result is consistent with the results of any other linear regression model in this research. It can be argued that it is not clear whether this model measures explicitly the effect of the proximity of parks with trees, or the effect of the proximity of a park in general as well. This would not be illogical since many academic studies show a beneficial effect of the accessibility of parks on the liveability of a certain place (see chapter 2).

In contrast to this model, a linear regression model with the independent variable "distance to the nearest park where 90% of the trees are high trees" does not result in a significant relationship (see appendices). It can be argued that less high trees in the park nearby is more preferred and that, therefore, some variety in vegetation – high as well as low vegetation like shrubs – might be preferred more.

The defects of Model 1 can be well visualized in the GIS-map below, Map 2. On the map, all valid observations (the green dots in Map 1) are represented and are colourized with different colours. The categories are divided through natural breaks. The map shows for each observation, how incorrect Model 1 estimated the relationship between the dependent variable (Y), the physical environment score, and independent variable (X1), the distance to a park with 70% high trees. The great variety of colours confirms the suspect that was suggested in the previous paragraph: Model 1 is not a good predictor of the relationship between Y and the distance to a park with many high trees.



Map 2: Residuals linear regression model 1: distance to the nearest park where 70% of the trees are high trees..

4.3 Model 2: Tree density

This section shows the results of the second linear regression model. As I have explained in chapter 3, three variables in the category 'tree density' are created. Due to the high tree density of 90%, which probably indicates a forest, only two variables are used in this research. Again, only one of them is significant. Table 5 presents the results of the model in which this significant variable is tested.

Linear Regression			**Number of observations**	2,204
			F (9,2194)	291.30
			Prob>F	0.0000
			R-squared	0.0850
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 70% of 0.5ha is forested	−.024451*	.011704	−.047403	−.001498
Standardized population density	−.122575*	.011331	−.14479	−.100353
Standardized property value	.056614*	.027634	.002423	.110806
Standardized percentage private houses	.241568*	.034172	.174555	.308582
Standardized percentage houses owned by a housing corporation	.101142*	.014646	.072419	.129864
Standardized number of thefts from houses and sheds	.027618*	.012128	.003833	.051403
Standardized number of acts of vandalism	.017739	.010712	−.003268	.038747
Standardized distance to the nearest family doctor	−.104247	.067702	−.237015	.028520
Standardized distance to the nearest supermarket	.066839	.052709	−.036525	.170204
Constant	.000000	.020416	−.040038	.040038

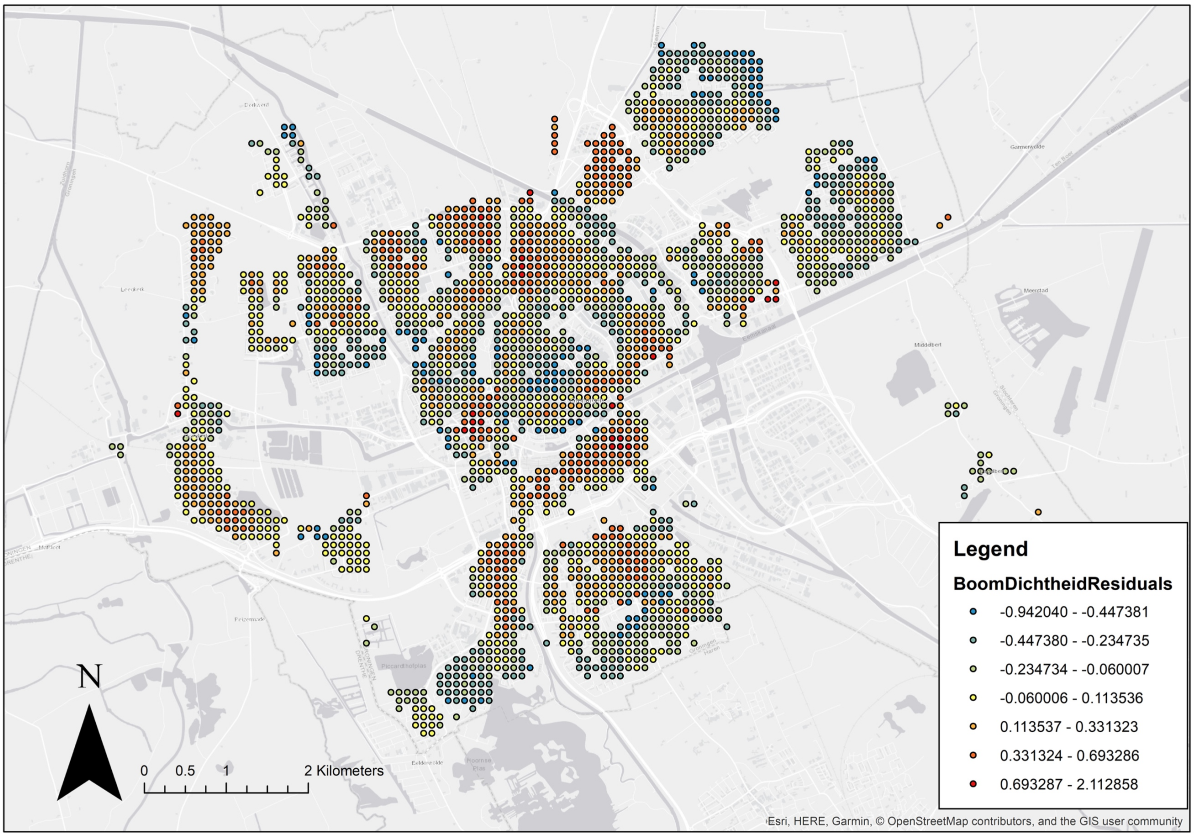
*Table 5: Model 2. A linear regression model with the dependent variable (Y), the physical environment score, and independent variable X2: distance to a park with a tree density of 70% (at least 70% of 0.5ha is forested). *P≤0.05.*

As shown above, the independent variable "distance to the nearest park where 70% of 0.5ha is forested" significantly affects the dependent variable "physical environment score" (p=0.037≤0.05). Besides that, the F-test (0.000) shows that the linear regression model itself is significant as well. These values are consistent with the results in the previous model, presented in Table 4. Also, the values are quite similar in the linear regression models with less or no control variables (see appendices).

It can be argued that the relationship between the dependent and the independent variable X2 is measured pretty accurate in this model. This accuracy appears from the comparison with another model (see Table 8, model 5 in the appendices). The independent variable "distance to a park with a tree density of 50%" (at least 50% of 0.5ha is forested), is not (however, almost) significantly related to the dependent variable ($p=0.082>0.05$). If we compare the robust standard errors of both models, we see a standard error of 0.011 (Model 2, X=distance to a park with a tree density of 70%) and a bigger standard error of 0.035 in the model with X=distance to a park with a tree density of 50%.

It is interesting to note that according to this model, the physical environment score declines with 0.024 standard deviation when a park with a tree density of 70% is located one standard deviation further away from the location of observation. As stated in the paragraph above, the independent variable "distance to a park with a tree density of 50%" does not have a significant relationship with the dependent variable. These results are consistent with the findings of the study of Bjerke et al. (2006), who mention that people prefer some vegetation density over an open scene.

Nevertheless, it can be argued that only eight percent of the variation in the dependent variable can be explained by the independent variable X2. Therefore, the same conclusion as for Model 1 applies: it is ambiguous what this model measures exactly. The differences between the observed and estimated values for each grid cell can again be visualised, which is presented in Map 3. The comments made on Map 2 (mentioned at the end of 4.2) apply to this map as well.



Map 3: Residuals linear regression model 2: distance to the nearest park where 70% of 0.5ha of 1ha is forested.

It can be argued that Model 2, the model with the independent variable "distance to the nearest park where 70% of 0.5ha is forested" seems to predict the dependent variable better than Model 1. In other words, there are fewer differences between the observed values and the predicted values in Model 2. A comparison of the robust standard errors of both models results in this suggestion as well: 0.0198 in Model 1 against 0.0117 in Model 2.

It is also interesting to note that the different categories of residuals seem to cluster more in Map 3 (Model 2, tree density) than in Map 2 (Model 3, tree height). It is natural that the wrong estimates cluster together, as a result of the First Law of Geography (Tobler, 1970). In my opinion, this could be the result of the fact that tree density seems to be more striking to residents, than tree height. After all, it is easier to determine how dense forest is, than to determine what the percentage 'high trees' in the forest is.

5. Conclusions, limitations, and discussion

In this study the question "*To what extent is the subjective quality of the physical environment in Groningen determined by the proximity of trees in the urban green space?"* was central. It was formulated in response to the increasing focus of Dutch municipalities on green spaces, and particularly trees, as a way to improve the physical environment. In Groningen, the proximity of a park with trees indeed positively affects the subjective quality of the physical environment.

Two of the four independent variables show a significant positive effect on the dependent variable: "the physical environment score" of the Dutch research, the *Leefbaarometer 2.0* (Ministry of the Interior and Kingdom Relations, 2016). It concerns the variables "Distance to nearest park where 70% of the trees are high trees" and "Distance to nearest park where 70% of 0.5ha is forested". In other words, as one moves further away from the nearest park with 70% high trees or with a tree density of 70%, the quality of the environment deteriorates.

These findings correspond with former research on the beneficial influence of green spaces on QoL and liveability. However, academic literature on the specific impact of trees is scarce. Mitchell (2013) states that physical activity in a forest is better for mental health than exercising in a non-natural environment. Furthermore, Kardan (2015) suggests physical health can be improved by the proximity of trees as well. Lastly, Franco & Macdonald (2018) argue that people are willing to pay a higher price for their house when it is located near a forest.

My research corresponds to the aforementioned studies and also offers a new perspective. The previous researchers studied small components of liveability like health and housing prices, also called 'indicators' (Ministry of the Interior and Kingdom Relations, 2016). In contrast, I studied the bigger picture, the dimension 'physical environment'. With this, my research contributes to the scarce academic knowledge regarding trees and liveability.

Apart from that, this research also has societal value. As stated in the introduction, Groningen has the goal to improve its living environment by planting trees and maintaining green spaces. I show that planting trees in parks will indeed improve Groningen's living environment. This knowledge can be used in policy-debates with respect to the ways to accomplish the coalition agreement goals.

Despite the fact that there is a significant and positive relationship between parks with trees and the subjective quality of the physical environment, there are also signs that parks with trees are not the determining factor. Firstly, at most nine percent of the variation in the dependent variable can be explained by the independent variables, which is very low. Secondly, often some of the control variables are significantly related to the dependent variable as well. This is not illogical because environmental elements always influence each other (Tobler, 1970).

Naturally, this research leaves much room for improvement. First of all, the independent variables seemed to be ambiguous. It is not clear if the positive impact on the dependent variable is specifically related to parks with trees, or if the effect of parks in general is measured. Further research could therefore try to design the research in such way that the effect of trees becomes clear. Below, I will suggest two options for this.

First, for this research certain percentages are chosen as requirements of independent variables. The results show that regarding tree height, the 70% variable is significant and the 90% variable is not. It can be argued that people prefer more vegetation variety in parks. It can be argued that they do not only want high trees, but lower trees and shrubs as well. With regards to the tree density, the 50% variable is not significant while the 70% variable is. Therefore, it can be argued that people prefer more density of vegetation in a park, which corresponds to the study of Bjerke et al. (2006). Between those percentages, there must be some turning point from significant to insignificant. Further research could attempt to discover this turning point.

Furthermore, different independent variables could be created and tested against the dependent variable. In this research, the tree height and density are the requirements and the variables are expressed in distances in meters. This could be done the other way around, by using the distance as requirement. In doing so, the height and density can vary. Consequently, knowledge about the influence of the height and density of trees in parks can be acquired. For every observation point can be determined whether a park with high or many trees is located, for instance, maximally 200 meters from the observation point.

Finally, more depth could be created in several ways. In addition to secondary data analysis, primary data could be collected and analysed. Besides that a qualitative element like in-depth interviews could be added to the research design. Also, a GIS-analysis could provide more depth seeing as it can provide insight in the types of areas where the model does not estimate the values accurately. Altogether, the outcomes of this study are very useful grounds for further research. Green spaces prove to be beneficial for people in numerous ways, and trees can certainly contribute to this. Groningen's intention to plant one hundred trees every year is therefore something we should definitely encourage.

References

Literature

Ambrey, C. & C. Fleming (2014). 'Public Greenspace and Life Satisfaction in Urban Australia', in *UrbanStudies* 51(6), 1290-1321.

Bertram, C. & K. Rehdanz (2015). 'The role of urban green space for human well-being', in *Ecological Economics* 120, 139-152.

Bjerke, T., Østdahl, T., Thrane, C. & Strumse, E. (2006). 'Vegetation density of urban parks and perceived appropriateness for recreation', in *Urban Forestry & Urban Greening* 5, 35-44.

Daams, M., Sijtsma, F. & A. van der Vlist (2016). 'The Effect of Natural Space on Nearby Propery Prices: Accounting for Perceived Attractiveness', in *Land Economics,* 92(3), 389-410.

Dzhambov, A., Markevych, I., Hartig, T., Tilov, B., Arabadzhiev, Z., Stoyanov, D., Gatseva, P. & D. Dimitrova (2018). 'Multiple pathways link urban green- and bluespace to mental health in young adults', in *Environmental Research* 166, 223-233.

Ekkel, E. & Vries, S. de (2017). 'Nearby green space and human health: Evaluating accessibility metrics', in *Landscape and Urban Planning* 157, 214-220.

Eurostat (2019). 'Quality of life indicators', *Eurostat. Statistics explained.* Available online: https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Quality_of_life_indicators (Accessed at: May 2, 2019)

Eurostat (2019a). 'Quality of life indicators – measuring quality of life', *Eurostat. Statistics explained.* Available online:
https://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_measuring_quality_of_life#B1_dimensions_of_quality_of_life

European Union (2019). Environment. Towards a greener and more sustainable Europe. Available online: https://europa.eu/european-union/topics/environment_en (Accessed at: May 3, 2019).

Franco, S. & J. Macdonald (2018). 'Measurement and valuation of urban greenness: Remote sensing and hedonic applications to Lisbon, Portugal', in *Regional Science and Urban Economics* 72, 156-180.

Gemeente Amsterdam (2018). *Een nieuwe lente, een nieuw geluid. Coalitieakkoord.* Available online: https://www.amsterdam.nl/bestuur-organisatie/volg-beleid/coalitieakkoord-2018/ (Accessed at: May 1, 2019)

Gemeente Den Haag (2018). *Den Haag: Stad van Kansen en Ambities. Coalitieakkoord 2018-2022.* Available online:
https://www.denhaag.nl/nl/bestuur-en-organisatie/college-van-burgemeester-en-wethouders/coalitieakkoord-2018-2022.htm (Accessed at: May 1, 2019)

Gemeente Eindhoven (2018). *Evenwicht & Energie. Zuinig op elkaar, op de stad en op de toekomst. Coalitieakkoord 2018-2022.* Available online:
https://www.eindhoven.nl/sites/default/files/2018-05/Coalitie%20magazine_0.pdf
(Accessed at: May 1, 2019)

Gemeente Groningen (2014). *Coalitieakkoord 2014-2018. Voor de Verandering.* Available online:
https://gemeente.groningen.nl/bestand/coalitieakkoord-2014-2018-voor-de-verandering
(Accessed at: May 1, 2019)

Gemeente Groningen (2019). *Coalitieakkoord 2019-2022. Gezond, groen, gelukkig Groningen.* Available online: https://gemeente.groningen.nl/coalitieakkoord (Accessed at: May 1, 2019)

Gemeente Nijmegen (2018). *Nijmegen: samen vooruit. Duurzaam, sociaal, veerkrachtig, aantrekkelijk. Coalitieakkoord 2018-2022.* Available online:
https://www.nijmegen.nl/fileadmin/bestanden/bestuur/college-B-W/Coalitieakkoord_2018-2022__Nijmegen_samen_vooruit.pdf (Accessed at: May 1, 2019)

Gemeente Rotterdam (2018). *Nieuwe energie voor Rotterdam. Coalitieakkoord 2018-2022.* Available online:
https://www.rotterdam.nl/bestuur-organisatie/college-van-benw/Coalitieakkoord-2018-2022.pdf (Accessed at: May 1, 2019)

Gemeente Utrecht (2018). *Utrecht. Ruimte voor iedereen. Coalitieakkoord 2018-2022.* Available online:
https://www.utrecht.nl/bestuur-en-organisatie/college-van-b-en-w/coalitieakkoord-2018/
(Accessed at: May 1, 2019)

Huang, Z. & X. Du (2015). 'Assessment and determinants of residential satisfaction with public housing in Hangzhou, China', in *Habitat International,* 47, 218-230.

Kaplan, R. (2001). 'The nature of the view from home. Psychological Benefits', in *Environment and Behavior* 33(4), 507-542.

Kardan, O., Gozdyra, P., Misic, B., Moola, F., Palmer, L., Paus, T. & M. Berman (2015). 'Neighborhood greenspace and health in a large urban center', in *Www.nature.com/scientificreports,* 5:11610, 1-14.
Available online:
https://www.nature.com/articles/srep11610 (Accessed at: May 21, 2019).

Klok, L., Rood, N., Kluck, J. & L. Kleereklopper (2019). 'Assessment of thermally comfortable urban spaces in Amsterdam during hot summer days', in *International Journal of Biometeorology,* 63, 129-141.

Kondo, M., Fluehr, J., McKeon, T. & C. Branas (2018). 'Urban Green Space and Its Impact on Human Health', in *International Journal of Environmental Research and Public Health* 15 (445), 1-28.

Leidelmeijer, K. & I. van Kamp (2003). *Kwaliteit van de Leefomgeving en Leefbaarheid. Naar een begrippenkader en conceptuele inkadering.* Amsterdam & Bilthoven: RIGO Research en Advies BV; RIVM.

MacKerron, G. & S. Mourato (2013). 'Happiness is greater in natural environments', in *Global Environmental Change* 23, 992-1000.
Mitchell, R. (2013). 'Is physical activity in natural environments better for mental health than physical activity in other environments?', in *Social Science & Medicine* 91, 130-134.

Mitchell, R. (2013). 'Is physical activity in natural environments better for mental health than physical activity in other environments?', in *Social Science & Medicine,* 91, 130-134.

Panduro, T., Jensen, C., Lundhede, T., Graevenitz, K. & B. Thorsen (2018). 'Eliciting preferences for urban parks', in *Regional Science and Urban Economics* 73, 127-142.

Punch, K. (2014). 'Controlling variables', in *Introduction to Social Research. Quantitative & Qualitative Approaches.* Thousand Oaks: Sage Publications Ltd, 221-223.

Sugiyama, T., Carver, A., Javad Koohsari, M. & J. Veitch (2018). 'Advantages of public green spaces in enhancing population health', in *Landscape and Urban Planning* 178, 12-17.

Susca, T., Gaffin, S., & G.R. Dell'Osso (2011). 'Positive effects of vegetation: Urban heat island and green roofs', in *Environmental Pollution,* 159, 2119-2126.

Tobler, W. (1970). 'A Computer Movie Simulating Urban Growth in the Detroit Region', in *Economic Geography,* 46, 234-240.

Vries, S. de, Verheij, R., Groenewegen, P. & P. Spreeuwenberg (2003). 'Natural environments – healthy environments? An exploratory analysis of the relationship between greenspace and health', in *Environment and Planning* 35, 1717-1731.

WHO Regional Office for Europe (2016). *Urban Green Spaces and Health.* Copenhagen.

Datasets and their appendices

Leefbaarometer 2.0

Ministry of Interior and Kingdom Relations (2018, December 12). *Leefbaarometer 2.0 = meting 2016* [Dataset]. Available online:
https://data.overheid.nl/data/dataset/leefbaarometer-2-0---meting-2016

RIGO Research & Atlas voor gemeenten (Leidelmeijer, K., Marlet, G., Ponds, R., Schulenberg, R., Woerkens, C. van & M. van Ham, 2014). *Leefbaarometer 2.0: Instrumentontwikkeling.* Available online:
https://doc.leefbaarometer.nl/resources/Leefbaarometer%202.0%20Instrumentontwikkelin g%20CONCEPT.pdf

Bestand Bodemgebruik/Document Soil Use

Central Statistical Office (2015). *Bestand Bodemgebruik* [Dataset]. Available online: https://data.overheid.nl/data/dataset/58880-bestand-bodemgebruik-2015

Central Statistical Office (2008). *Bestand Bodemgebruik. Productbeschrijving.* Available online: https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/bestand-bodemgebruik-bbg/classificaties/-/media/3a93f4f3f11648d4a6b9c3ff01864cc4.ashx

Bomen in Nederland/Trees in the Netherlands

Atlas Natuurlijk Kapitaal [ANK] (2017, October 6). *Bomen in Nederland* [Dataset]. Available online: http://www.atlasnatuurlijkkapitaal.nl

Kerncijfers wijken en buurten/Key figures neighbourhoods and areas

Central Statistical Office (2019b, March 29). Kerncijfers wijken en buurten [Dataset]. Available online: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84286NED/table

Central Statistical Office (2018). *Toelichting Wijk- en Buurtkaart 2016, 2017 en 2018.* Available online: https://www.cbs.nl/-/media/cbs/dossiers/nederland%20regionaal/wijk-en-buurtstatistieken/2018/2018ep42%20toelichting%20buurtkaart%202018.pdf

Ecosystem Unit Map

Central Statistical Office (2017a, March 24). Ecosystem Unit Map [Dataset]. Available online: https://www.cbs.nl/en-gb/background/2017/12/ecosystem-unit-map

Central Statistical Office (2017b). Ecosystem Unit Map, product description. Available online: https://www.cbs.nl/- /media/_pdf/2017/12/2017ep14%20ecosystem%20unit%20map.pdf

Other

Central Statistical Office (2019a). *Regionale kerncijfers Nederland. Regio's, 2019, totale bevolking* [Dataset]. Available online: https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70072ned

Acknowledgements

I want to thank prof. dr. D. Ballas for his supervision and good advice on regression models, the use of control variables and other issues regarding my research. This was my first statistical practice ever, and his advice helped me improving my academic skills.

My special and great gratitude goes to Richard Rijnks, MSc. He helped me to put my ideas into practice, and he invested many hours in merging the datasets for me and creating all interaction variables.

Furthermore, I want to thank Viktor Venhorst for answering my remaining statistical questions and Luuk Meijer for showing me how StataSE works.

Last but not least, my sincere gratitude to Rogier Meerlo, who sat with me for hours to create the GIS-maps used in this thesis.

Appendix A: Linear regression output

I. Distance to a park with X% of trees are high trees

Linear Regression			**Number of observations**	2,204
			F (1,2202)	8.35
			Prob>F	0.0039
			R-squared	0.0033
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 70% of the trees are high trees	−.057564*	.019916	−.096622	−.018506
Constant	.000000	.021270	−.041711	.041711

*Table 6: Model 3. A single linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park, of which 70% of the trees are high trees. *P≤0.05.*

Linear Regression			**Number of observations**	2,204
			F (10,2193)	205.43
			Prob>F	0.0000
			R-squared	0.0000
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 70% of the trees are high trees	−.044672*	.019504	−.082921	−.006422
Standardized population density	−.172145*	.010255	−.192257	−.152032
Standardized percentage private houses	.182292*	.023400	.136402	.228181
Standardized number of thefts from houses and sheds	.011793	.008216	−.004319	.027905
Standardized distance to the nearest family doctor	−.076032	.046076	−.166389	.014325
Constant	.000000	.020453	−.040038	.040038

*Table 7: Model 4. A multiple linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park, of which 70% of the trees are high trees. *P≤0.05.*

Linear Regression			**Number of observations**	2,204
			F (10,2193)	289.11
			Prob>F	0.0000
			R-squared	0.0844
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 90% of the trees are	−.000065	.006138	−.012102	.011972

high trees				
Standardized population density	−.124622*	.011431	−.147040	−.102203
Standardized property value	.055562*	.027991	.000670	.110453
Standardized percentage private houses	.241193*	.033856	.174800	.307586
Standardized percentage houses owned by a housing corporation	.098941*	.014794	.069929	.127954
Standardized number of thefts from houses and sheds	.026500*	.011684	.003587	.049413
Standardized number of acts of vandalism	.013541	.010660	−.007364	.034447
Standardized distance to the nearest family doctor	−.107456	.068453	−.241695	.026783
Standardized distance to the nearest supermarket	.067737	.052892	−.035987	.171462
Constant	.000000	.020423	−.040051	.040051

*Table 8: Model 5. A multiple linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park, of which 90% of the trees are high trees. *P≤0.05.*

II. Distance to a park with X% of 0.5ha is forested

Linear Regression			**Number of observations**	2,204
			F (10,2193)	286.07
			Prob>F	0.0000
			R-squared	0.0882

Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 50% of 0.5ha is forested	.061765	.035539	−.007929	.131460
Standardized population density	−.118986*	.011048	−.140652	−.097320
Standardized property value	.055515*	.027969	.000666	.110365
Standardized percentage private houses	.245189*	.035700	.175178	.315201
Standardized percentage houses owned by a housing corporation	.102740*	.014361	.074577	.130904
Standardized number of thefts from houses and sheds	.030343*	.013197	.004462	.056225

Standardized number of acts of vandalism	.010838	.011120	−.010970	.032646
Standardized distance to the nearest family doctor	−.107476	.068317	−.241450	.026497
Standardized distance to the nearest supermarket	.068092	.052837	−.035524	.171709
Constant	.000000	.020381	−.039968	.039968

*Table 9: Model 6. A multiple linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park with a tree density of 50% (at least 50% of 0.5 ha is forested). *P≤0.05.*

Linear Regression			**Number of observations**	2,204
			F (1,2202)	12.39
			Prob>F	0.0004
			R-squared	0.0027
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 70% of 0.5ha is forested	−.051543*	.014640	−.080254	−.022832
Constant	.000000	.021277	−.041725	.041722

*Table 10: Model 7. A single linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park with a tree density of 70% (at least 70% of 0.5 ha is forested). *P≤0.05.*

Linear Regression			**Number of observations**	2,204
			F (5,2198)	254.77
			Prob>F	0.0000
			R-squared	0.0786
Physical Environment Score	**Coef.**	**Robust Std. Error**	**[95% Conf. Interval]**	
Standardized distance to nearest park where 70% of the trees are high trees	−.021809	.011991	−.045324	.001705
Standardized population density	−.167754*	.010051	−.187466	−.148043
Standardized percentage private houses	.184548*	.024419	.136661	.232436
Standardized number of thefts from houses and sheds	.011037	.008055	−.004759	.026834
Standardized distance to the nearest family doctor	−.072834	.045189	−.161452	.015783
Constant	.000000	.020469	−.040142	.040142

*Table 11: Model 8. A multiple linear regression model with the dependent variable (Y), the physical environment score, and independent variable X: distance to a park with a tree density of 70% (at least 70% of 0.5 ha is forested). *P≤0.05.*

IV: Correlations

(obs=2,204)	zParkHGT70_Y	zDichtheid70
zParkHGT70_Y	1.0000	
zDichtheid70	0.4966	1.0000

Table 13: A correlation table with variables tree height 70% and tree density 70%.

Variable	VIF	1/VIF
zP_KOOPWON	7.78	0.128594
zP_HUURCORP	4.31	0.231948
zBEV_DICHTH	2.86	0.349831
zWOZ_v	2.17	0.460724
zAFARTSPR	2.15	0.464044
zAFSUPERM	2.09	0.478556
zvernieling	1.22	0.817249
zwoningdief	1.22	0.820316
zParkHGT90_Y	1.07	0.934126
zDichtheid50	1.04	0.962713
Mean VIF	2.59	

Table 14: A multicollinearity table. Conclusion: multicollinearity out of question.