Geography
Internal Assessment

How does environmental quality, determined by the severity of pollution and the availability of green spaces, differ between a tourist-catering area of Paris versus a residential area?

Väinö-Verneri Kauppila
May 2021

Word count: 2472

Contents

1 Introduction **1**

1.1 Fieldwork question 1

1.2 Hypotheses 1

2 Method **2**

2.1 Study site choices 2

2.1.1 Our sampling method 4

2.2 Method 4

2.2.1 Bipolar semantic survey 4

2.2.2 Statistical evaluations 4

2.2.3 Questionnaire 5

3 Data & Analysis **5**

3.1 Data collected 5

3.2 Analysis 8

3.2.1 Bipolar survey 8

3.2.2 Nearest Neighbor Index 9

3.2.3 Questionnaire 10

4 Conclusion **10**

4.1 Evaluation 11

Notes **i**

A Appendix **i**

A.1 Bipolar survey results i

A.2 NNI computation program ii

1 Introduction

1.1 Fieldwork question

How does environmental quality, determined by the severity of pollution and the availability of green spaces, differ between a touristic area of Paris and a residential area?

In order to answer this question, the fieldwork for our paper was conducted in Paris, one of the largest and most well known cities in the world. The city itself is arranged into 20 districts, known locally as *arrondissements*, are placed in a spiral in the city. It is globally known for its high number of tourists per year, equating to around 35 million in the year 2019 alone.[1] Many of them come to visit the world-renowned Eiffel Tower, located in the VII$^{\text{th}}$ *arrondissement*.[2] To add, Paris is home to 2.2 million people, who mostly live in the outer residential areas, from the XI$^{\text{th}}$ to the XX$^{\text{th}}$ *arrondissements*.

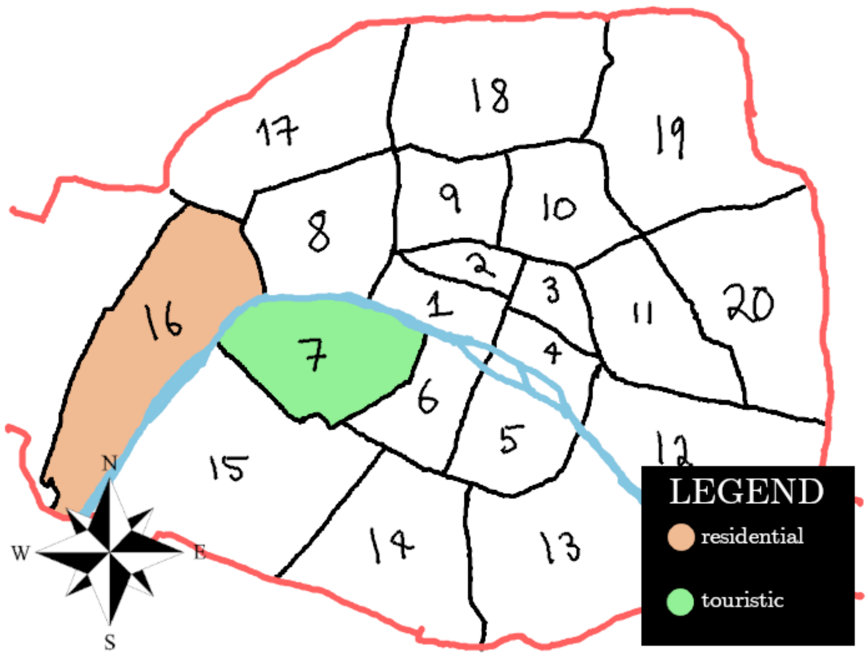

Figure 1: A map of Paris *arrondissements*, drawn by hand.

1.2 Hypotheses

1. According to the Global Development Goals of the UN, 90% of urban areas in the world had polluted air in 2016.[3] Paris was among countries that didn't satisfy WHO's air quality minimum[4] of 2018, with on average a 50% higher than normal pollution density. However, a

[1]Statista Research Department. *Hotel arrivals in Paris 2011-2019*. Apr. 2020. URL: https://www.statista.com/statistics/468164/number-tourist-arrivals-hotels-paris.

[2]CondorFerries. *Latest France Tourism Statistics & Industry Trends (2020-2021)*. URL: https://www.condorferries.co.uk/france-tourism-statistics.

[3]"The Sustainable Development Goals Report 2020". In: *The Sustainable Development Goals Report* (2020), p. 47. DOI: 10.18356/214e6642-en.

[4]*Ambient (outdoor) air pollution*. May 2018. URL: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

counterargument to this could be that despite a higher concentration of people, tourist areas in Paris do not suffer as much from high traffic conditions from things such as typical morning rush hours, and tourists preferably using public transport or bikes.

2. As there are more people moving about in residential areas, for example in cars for the morning commute or at noon for lunch, it can logically be theorized that noise pollution, which is obviously a function of the amount of cars, would be higher in these places. Today it is estimated that an average noise level of $60dB$ can be found in residential areas, according to the very comprehensive Bruitparif government-sponsored report.[5] This value largely surpasses the WHO's safe level of $53dB$.[6]

3. The Parisian mayor, Anne Hidalgo included the improvement of environmental quality in her campaign. The mayor has promised to make so-called "green spaces" no further than 200 meters to any person,[7] and as such, it should be hypothesized that green spaces, which include parks, agglomerations of trees, shall be distributed evenly with no difference between residential and touristic areas. The mayor emphasized on "urban forests" — places where residents and tourists alike could enjoy the company of trees while walking along the city streets.

2 Method

For our investigation, the topic in question is the environmental quality. We will compare the environmental quality of two areas, one meant as a residential one and one with a heavy tourist presence. To best represent these areas, we have chosen the XVIe and the VIIe. The XVIe is home to many housing complexes and fosters facilities aimed at catering to the residents whereas the VIIe sees many tourists as it is home to the famed Eiffel Tower and the Seine river, prime tourist attractions of Paris.

2.1 Study site choices

We chose 10 sites in total to conduct a bipolar survey, shown below in Figures 2, 3.

[5] Mairie de Paris. *PLAN DE PRÉVENTION DU BRUIT DANS L'ENVIRONNEMENT 2015 > 2020*. URL: https://www.bruitparif.fr/PPBE/75056%20-%20Paris/PPBE%20Paris%202015-2020.pdf.

[6] "Environmental Noise Guidelines for the European Region". In: (2018), p. 8. URL: https://www.euro.who.int/__data/assets/pdf_file/0009/383922/noise-guidelines-exec-sum-eng.pdf.

[7] Anne Hidalgo. *Comment Paris peut-elle être une ville encore plus végétale ?* URL: https://annehidalgo2020.com/question/comment-paris-peut-etre-une-ville-encore-plus-vegetale/.

Figure 2: Map of sites chosen for the residential area, the XVIth *arrondissement*. Map base layer courtesy of Google Maps, pictures seen are taken on a mobile phone.

This site was chosen as, like stated before, it is a residential area. There are other residential areas in Paris, however this one was specifically chosen for reasons of convenience.[8]

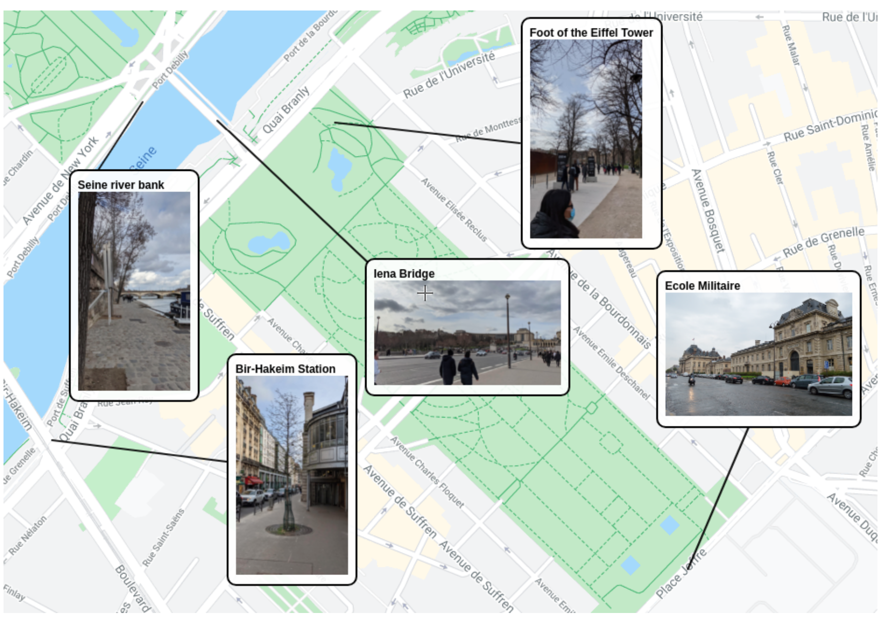

Figure 3: Map of sites chosen for the touristic area, the VIIth *arrondissement*. Map base layer courtesy of Google Maps, pictures seen are taken on a mobile phone.

[8] We had 2 members of our group residing in this area.

This area in the VIIth was chosen as it is a staple of tourism in France. It is home to the famed Eiffel Tower, the Ecole Militaire building and the Seine river bank.

2.1.1 Our sampling method

The sites in the two areas were chosen using a method of stratified sampling. Considering the VIIth does indeed contain residential sites and the XVIth touristic ones, we can discard the ones that are irrelevant to our study.

2.2 Method

2.2.1 Bipolar semantic survey

The sites were visited to conduct a so-called "bipolar survey", which consisted of rating the site based on multiple criteria. This process was done by selecting a single person from our group to visit one of the sites, and complete a bipolar semantic survey. They would grade several different aspects of the site (such as general cleanliness, noise, amount of cars), and take a set of 3 images of the site.[9] Our surveys were conducted during the week of the 10th of March, 2021 in the afternoon each time. The weather was overcast and moderately cold.

2.2.2 Statistical evaluations

In order to either support or disprove our third hypothesis, it is important to study how trees are scattered in Paris, which reveals how dispersed green spaces are from one another. A statistical test called the Nearest Neighbor Index (NNI) test can provide a reasonable answer to this: if the index is similar for both sites, it can easily be said then that our hypothesis is supported. This value also reveals information about the distribution of trees: clustered, random or in a regular pattern, as seen in Figure 4 below. These values will be presented and analysed in the later section 3. Data & Analysis.

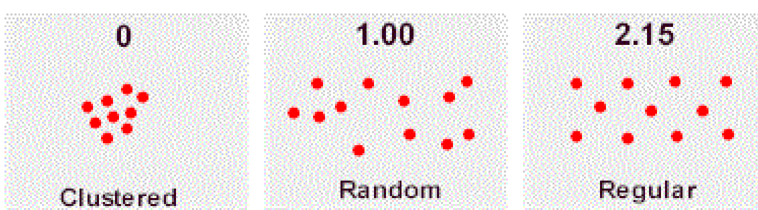

Figure 4: The NNI measures the spacial distribution of data. From a value of 0, representing a clustered pattern, to 1 (random) and to 2.15 (uniform pattern)

As individually counting trees in these areas would be a tedious task, we used the Paris OpenData platform (licensed under the permissive ODbL license[10]), owned by the government. The fact that

[9] Images taken for our study can be found at `https://github.com/kinnounko/notes/tree/main/geography/ia/images`

[10] *Open Data Commons Open Database License (ODbL) - Open Data Commons: legal tools for open data.* URL: `https://opendatacommons.org/licenses/odbl/`.

this kind of database platform exists, shows a certain level of transparency on the government's part.

We used the trees dataset procured from the library, which contains the exact coordinates of trees in the city.[11] There are also other information such as species of the tree and such, however these are cleaned from the dataset for our purposes.

A script, written in Python was tasked to calculate the NNI value, because with a dataset of over 20,000 data points this would not be possible by hand. We go through each point, and find its nearest neighbor, using an efficient complex data structure called a kd-tree.[12] In order to calculate distances between these two points, we use the Haversine Formula, as the two points would be in the form of coordinates. With this information the NNI can be calculated with the simple formula

$$Rn = \frac{2\bar{D}}{\sqrt{\frac{a}{n}}}$$

where Rn is the NNI index value, $\bar{D}$ is the mean observed distance to the nearest neighbor, a is the area of the zone and n is the total number of data points.

2.2.3 Questionnaire

We also collected data from a questionnaire sent out to many people. This questionnaire contained the two questions *"Do you believe that the Passy area has more green spaces than the area around the Eiffel tower? (Not including the champ de mars)"* and *"How much litter is in the Passy area compared to around the Eiffel tower?"*. The data collected from this survey is listed in Section 3.1. We surveyed our year, which has people that live in and frequent these two areas.

3 Data & Analysis

Once our data collected, we can analyse certain patterns or appearances that do not follow a general trend. Our data is shown below in Section 3.1.

3.1 Data collected

Our results for the bipolar survey are shown in Appendix A.1 (not shown here due to large size). When tallied, the total scores for each site can be expressed as a fraction over 84, as shown in Figures 5, 6.

[11] *Les arbres.* May 2021. URL: https://opendata.paris.fr/explore/dataset/les-arbres/information.

[12] k-dimensional tree. In this case, this "tree" is 2-dimensional. More information at: Jon Louis Bentley. "Multidimensional Binary Search Trees Used for Associative Searching". In: *Stanford University* (1975), pp. 1–9. URL: http://web.mit.edu/bookwyrm/Public/p509-bentley.pdf

Site	Total score
Bir-Hakeim	51
Seine river bank	69
Foot of the Eiffel Tower	45
Iena Bridge	42
Ecole Militaire	54

Figure 5: Total Bipolar Semantic scores for the VII$^{\text{th}}$

Site	Total score
Rue Davioud 1	47
Rue Davioud 2	53
Avenue Mozart	31
Avenue Paul Doumer	35
Ranelagh Garden	77

Figure 6: Total Bipolar Semantic scores for the XVI$^{\text{th}}$

When plotted on a map, these totals can be shown as percentages ($\frac{s}{84} \cdot 100$, where s is the total score) with a gradient of colors, as seen below in Figures 7, 8:

Figure 7: Each site in the VVI$^{\text{th}}$, with its respective total score expressed as a percentage for clarity, and a color assigned based on the gradient in Figure 9.

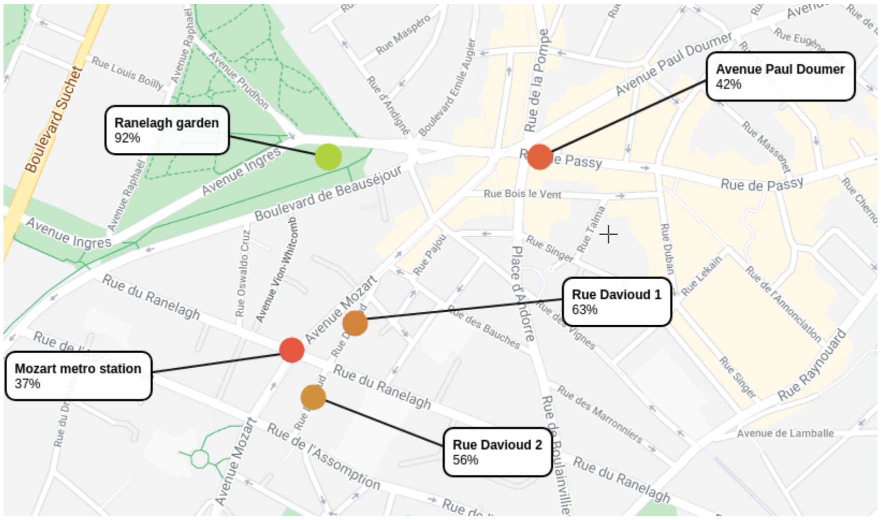

Figure 8: Each site in the XVI$^{\text{th}}$, with a percentage. Color assigned with gradient in Figure 9.

Figure 9: Gradient from red (hex color #FF0000) for 0%, to green (hex color #00FF00) for 100%.

As for images collected during our inspections of the sites, for the sake of keeping the file size of this paper down, I will not attach the images collected and they will be available for download or viewing at the URL `https://github.com/kinnounko/notes/tree/main/geography/ia`.

The dataset collected from the questionnaire (mentioned in Section 2.2.3) is shown in the tables in Figures 10, 11.

Answer	Number of people that agree
Significantly Less	4
Less	8
About the same	8
More	10
Significantly more	0

Figure 10: *Do you believe that the Passy area has more green spaces than the area around the Eiffel tower? (Not including the champ de mars)*

Answer	Number of people that agree
Significantly Less	3
Less	10
About the same	12
More	5
Significantly more	0

Figure 11: *How much litter is in the Passy area compared to around the Eiffel tower?*

This data can be shown in the form of a bar graph (Figure 12), in order to visualize the distribution and for further analysis.

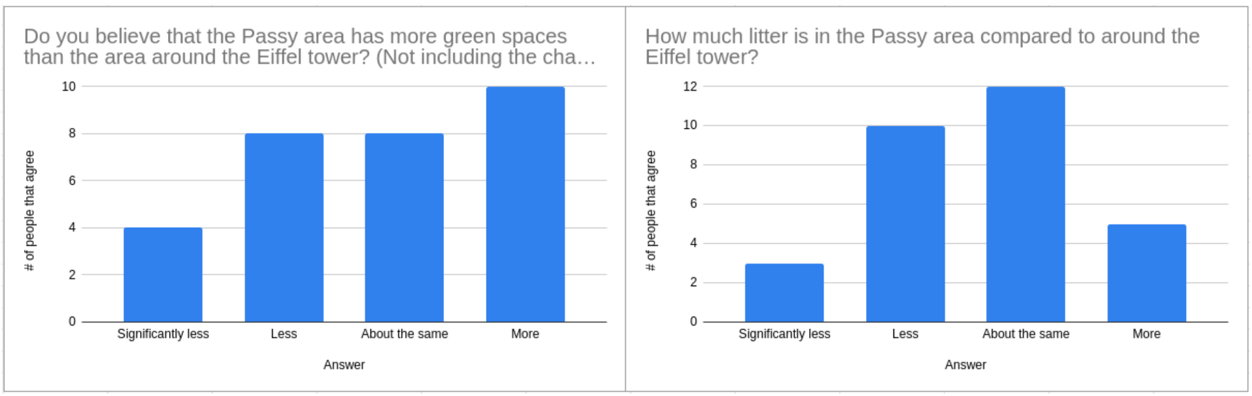

Figure 12: Graphs representing aforementioned data in Figures 10, 11

3.2 Analysis

3.2.1 Bipolar survey

The results from our bipolar semantic survey seem to suggest a few things. Firstly, for the VIIth, the mean of the data series seems to be an overall environmental percentage value of 62.5%. We can notice an outlier in this data: the Seine river bank, with a score of 82%. This higher score is due to the nature of this site: no cars, low vandalism and no odor (due to its proximity to a body of water). In general however, it seems that the more tourist-heavy areas are more prone to lower scores, such as the foot of the Eiffel Tower and the Iena Bridge.

In the XVIth, the results differ from the ones in the touristic area. Again, the mean of the data series is 58%, considerably lower than that of the touristic area. This does include the obvious outlier however, the Ranelagh garden at 92%. Considering the sites with higher car counts and noise levels (such as Avenue Mozart) are in this residential area, this does indeed agree with our 2nd hypothesis.

In general, it can be said with this information that statistically, our bipolar semantic survey agrees with the 1st and 2nd hypotheses made at the beginning of our paper: the residential area does indeed show higher levels of noise and a lower environmental score.

3.2.2 Nearest Neighbor Index

We must analyse the NNI values from the Nearest Neighbor statistical test in order to either prove or disprove the 3rd hypothesis we made. With the `arbres` dataset downloaded, we can use the `nearest_neighbor_index.py` [13] program in order to calculate the NNI values of the XVIth and the VIIth. We must use the following syntax:

```
python nearest_neighbor_index.py -b x1,y1,x2,y2 /folder/dataset.csv
```

Or in practice, for the XVIth:

```
python nearest_neighbor_index.py -b
    48.85373,2.26728,48.86042,2.27865 /home/octo/trees.csv
```

And for the VIIth:

```
python nearest_neighbor_index.py -b
    48.84881,2.28708,48.86075,2.30742 /home/octo/trees.csv
```

This will output the NNI values for the XVIth and the VIIth, respectively. They are shown in the table below in Figure 13.

Area	NNI value (no unit, rounded to 3 s.f.)
XVIth	0.610
VIIth	0.534

Figure 13: Computed NNI values

This information allows us to reach a conclusion about the similarity of the dispersion of trees in the two areas: both values are between the "Clustered" value of 0.00 and the "Random" value of 1.00, and as such are both dispersed spatially in a clustered-random manner. We are however not interested in their manner of dispersion but the similarity between the two values. And as it can be seen, these values are very similar.

In fact, if we observe the graph below in Figure 14, we can see that our NNI values express very "significant element of clustering", as our datasets have a number of points of 1277 and 5231 (values that are not included in the x-axis of this graph due to their large size):

[13] The program can be found at `https://github.com/kinnounko/notes/blob/main/geography/ia/nearest_neighbor_index.py` or Appendix A.2

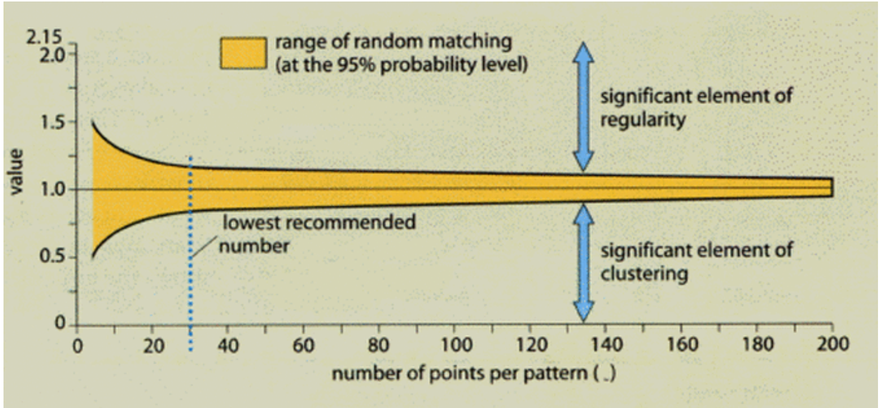

Figure 14: A graph showing the certainty of a NNI value result

And as such, to conclude, the 3rd hypothesis of our research is confirmed (at least for the two study areas). We can with confidence say that trees, and as a result green spaces, are evenly spread in the two areas. They also exhibit signs of clustering, which signifies the presence of these trees in groups, or parks. As the mayor has stated, it does seem that green spaces are available to everyone.

3.2.3 Questionnaire

As seen in Figure 12, our questionnaire brought some mixed opinions. Many believed that the XVIth has more green spaces surrounding it, as shown by the negatively skewed bar chart. People also believed that the residential area in the VIIth was less littered than the touristic one, as shown by again a negatively skewed graph.

This information does go against our hypotheses, but as this is an opinionated survey and because I believe some people have not visited one of the sites, I will not allocate much weight to these results and rather rely on empirical evidence.

4 Conclusion

With this information analysed, we can reach a sensible conclusion. Paris as a city is generally not very favorable environmentally: with high air and noise pollution in residential areas and other issues, the most habited places of the city are not very nice to be in. However touristic areas do show a small but noticeably higher level of cleanliness.

There is however a visible effort in order to fix some of these more unfavorable areas. It could be for press-related issues, with the upcoming 2024 Paris Olympics, or by pressure from international organizations with their pollution guidelines, but there is a noticeable effort being done: promises of parks for everyone by the mayor seem to have been fulfilled and are valid, and more regulations are being put into place as evidenced by the 88-page manifesto-like government sponsored brochure.[14] These improvements are not uniform in the city, and personally they seem to be focused in touristic and publicized areas. Throughout my time in Paris, I have seen many improvements to environmental quality, albeit predominately in these touristic areas, and zero to no effort in residential areas such

[14] Paris, *PLAN DE PRÉVENTION DU BRUIT DANS L'ENVIRONNEMENT 2015 > 2020*.

as the one I live in. In order to promote environmental sustainability in this city, there must be some action taken in residential areas, and not only in tourist-frequented ones.

4.1 Evaluation

There are some processes in our paper that would be done differently if done now. To list the ones I have managed to find at the time of writing:

- The computer program does not seem to take into account an area of greenery in the VII$^{\text{th}}$, which can change the NNI value found for the VII$^{\text{th}}$. This area includes parks and such, however the map (generated also with the computer program) does not show trees in this location. This should have been investigated, however due to constraints related to time, I could not. I suspect this was due to a preliminary cleanup of the data. I do not think this error has affected the outcome of our research by a lot, as with these parks the value would be around the same, as they also show signs of clustering.

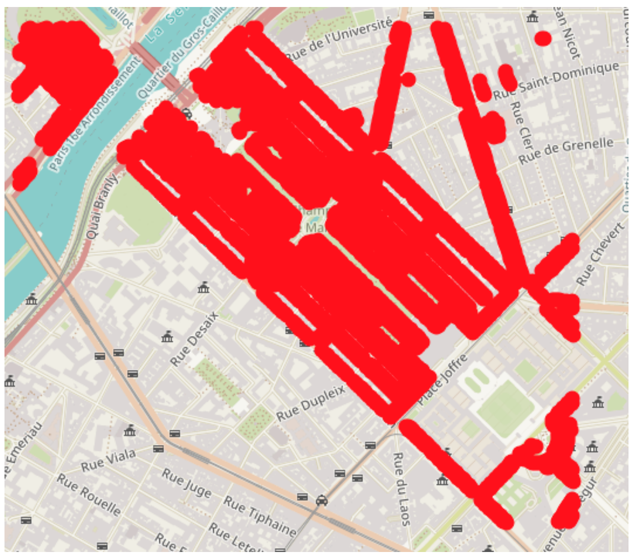

Figure 15: A map showing the trees in the dataset used for the VII$^{\text{th}}$, along with some parks in the bottom left quadrant, missing their trees.

- Our research was done during a time of lock-down due to the SARS-COVID-19 pandemic in France. Due to travel restrictions, international — even national tourist counts were considerably down. This causes many discrepancies in the data as, well, we are in fact studying a touristic area.

To conclude the evaluation, it can be said that despite some disparity in our data, our experimental method was correct, computational issues aside. Once the SARS-COVID-19 pandemic over, this research could be repeated once again for more accurate results.

Notes

This paper is written with the aid of the typesetting software LaTeX. On a PDF viewer, the sections in the table of contents can be clicked to access that section.

All media and information related to this IA can be found on the URL https://github.com/kinnounko/notes/tree/main/geography/ia. This includes images, the LaTeX source code to this paper, and the Python program used to calculate the NNI value (also found in Appendix A.2).

References

Department, Statista Research. *Hotel arrivals in Paris 2011-2019*. Apr. 2020. URL: https://www.statista.com/statistics/468164/number-tourist-arrivals-hotels-paris.

CondorFerries. *Latest France Tourism Statistics & Industry Trends (2020-2021)*. URL: https://www.condorferries.co.uk/france-tourism-statistics.

"The Sustainable Development Goals Report 2020". In: *The Sustainable Development Goals Report* (2020), p. 47. DOI: 10.18356/214e6642-en.

Ambient (outdoor) air pollution. May 2018. URL: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

Paris, Mairie de. *PLAN DE PRÉVENTION DU BRUIT DANS L'ENVIRONNEMENT 2015 > 2020*. URL: https://www.bruitparif.fr/PPBE/75056%20-%20Paris/PPBE%20Paris%202015-2020.pdf.

"Environmental Noise Guidelines for the European Region". In: (2018), p. 8. URL: https://www.euro.who.int/__data/assets/pdf_file/0009/383922/noise-guidelines-exec-sum-eng.pdf.

Hidalgo, Anne. *Comment Paris peut-elle être une ville encore plus végétale ?* URL: https://annehidalgo2020.com/question/comment-paris-peut-etre-une-ville-encore-plus-vegetale/.

Open Data Commons Open Database License (ODbL) - Open Data Commons: legal tools for open data. URL: https://opendatacommons.org/licenses/odbl/.

Les arbres. May 2021. URL: https://opendata.paris.fr/explore/dataset/les-arbres/information.

Bentley, Jon Louis. "Multidimensional Binary Search Trees Used for Associative Searching". In: *Stanford University* (1975), pp. 1–9. URL: http://web.mit.edu/bookwyrm/Public/p509-bentley.pdf.

A Appendix

A.1 Bipolar survey results

Shown below are our results from our bipolar semantic survey:

XVI$^{\text{th}}$ residential area:

AVENUE MOZART

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism			X					No vandalism
Noisy				X				Calm
Crowded			X					Spacious
Lots of cars	X							No cars
Lots of smokers					X			No smokers
Lots of litter		X						No litter
Poorly maintained				X				Well maintained
No nature	X							Lots of nature
Overfilled bins		X						Empty bins
No recycling bins		X						Lots of recycling bins
Foul / polluted air				X				No noticeable odor
Lots of dog feces		X						No dog feces
								Total: 31/84

RUE DAVIOUD

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism			X					No vandalism
Noisy						X		Calm
Crowded						X		Spacious
Lots of cars					X			No cars
Lots of smokers						X		No smokers
Lots of litter				X				No litter
Poorly maintained					X			Well maintained
No nature								Lots of nature
Overfilled bins						X		Empty bins
No recycling bins					X			Lots of recycling bins
Foul / polluted air				X				No noticeable odor
Lots of dog feces			X					No dog feces
								Total: 47/84

RUE DAVOIUD 2

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism						X		No vandalism
Noisy	/	/	/	/	X	/	/	Calm
Crowded					X			Spacious
Lots of cars						X		No cars
Lots of smo	X							No smokers
Lots of litter							X	No litter
Poorly maintained						X		Well maintained
No nature		X						Lots of nature
Overfilled bins				X				Empty bins
No recycling bins			X					Lots of recycling bins
Foul / polluted air						X		No noticeable odor
Lots of dog feces			X					No dog feces
								Total : 53/84

RANELAGH JARDIN

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism							X	No vandalism
Noisy						X		Calm
Crowded						X		Spacious
Lots of cars							X	No cars
Lots of smokers							X	No smokers
Lots of litter						X		No litter
Poorly maintained							X	Well maintained
No nature							X	Lots of nature
Overfilled bins						X		Empty bins
No recycling bins						X		Lots of recycling bins
Foul / polluted air							X	No noticeable odor
Lots of dog feces					X			No dog feces
								Total: 77/84

AVENUE PAUL DOUMER

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism		X						No vandalism
Noisy		X						Calm
Crowded		X						Spacious
Lots of cars	X							No cars
Lots of smokers			X					No smokers
Lots of litter				X				No litter
Poorly maintained						X		Well maintained
No nature		X						Lots of nature
Overfilled bins			X					Empty bins
No recycling bins				X				Lots of recycling bins
Foul / polluted air				X				No noticeable odor
Lots of dog feces			X					No dog feces
								Total: 35/84

VIIth touristic area:

BIR HAKEIM METRO STATION

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism						X		No vandalism
Noisy		X						Calm
Crowded					X			Spacious
Lots of cars				X				No cars
Lots of smokers					X			No smokers
Lots of litter			X					No litter
Poorly maintained			X					Well maintained
No nature			X					Lots of nature
Overfilled bins				X				Empty bins
No recycling bins					X			Lots of recycling bins
Foul / polluted air					X			No noticeable odor
Lots of dog feces						X		No dog feces
								Total: 51/84

ECOLE MILITAIRE

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism			X					No vandalism
Noisy					X			Calm
Crowded						X		Spacious
Lots of cars			X					No cars
Lots of smokers			X					No smokers
Lots of litter				X				No litter
Poorly maintained					X			Well maintained
No nature					X			Lots of nature
Overfilled bins						X		Empty bins
No recycling bins					X			Lots of recycling bins
Foul / polluted air				X				No noticeable odor
Lots of dog feces					X			No dog feces
								Total: 54/84

FOOT OF THE EIFFEL TOWER

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vanc	X							No vandalism
Noisy				X				Calm
Crowded		X						Spacious
Lots of cars				X				No cars
Lots of smokers			X					No smokers
Lots of litter			X					No litter
Poorly maintained						X		Well maintained
No nature						X		Lots of nature
Overfilled bins					X			Empty bins
No recycling bins			X					Lots of recycling bins
Foul / polluted air						X		No noticeable odor
Lots of dog feces			X					No dog feces
								Total: 45/84

PONT D'IÉNA

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism							X	No vandalism
Noisy					X			Calm
Crowded	X							Spacious
Lots of cars		X						No cars
Lots of smokers		X						No smokers
Lots of litter			X					No litter
Poorly maintained						X		Well maintained
No nature		X						Lots of nature
Overfilled bins			X					Empty bins
No recycling	X							Lots of recycling bins
Foul / polluted air					X			No noticeable odor
Lots of dog feces						X		No dog feces
								Total: 42/84

BANK OF THE SEINE

NEGATIVE	1	2	3	4	5	6	7	POSITIVE
Lots of vandalism						X		No vandalism
Noisy							X	Calm
Crowded						X		Spacious
Lots of cars							X	No cars
Lots of smokers						X		No smokers
Lots of litter				X				No litter
Poorly maintained					X			Well maintained
No nature				X				Lots of nature
Overfilled bins						X		Empty bins
No recycling bins					X			Lots of recycling bins
Foul / polluted air							X	No noticeable odor
Lots of dog feces						X		No dog feces
								Total: 69/84

A.2 NNI computation program

```
import pandas as pd
import math
from scipy.spatial import KDTree
import numpy as np
from statistics import mean
import argparse


def box_select_coordinates(df, bndg_box):
    # Box select a certain region of geographical coordinates from a df
    x_bound = df[(bndg_box[0][0] < df['lat']) & (bndg_box[1][0] > df['
        lat'])]
    return x_bound[(bndg_box[0][1] < df['lon']) &
                   (bndg_box[1][1] > df['lon'])]


def box_area(bndg_box):
    dx = haversine_dist(bndg_box[0][0], bndg_box[0]
                        [1], bndg_box[1][0], bndg_box[0][1]) * 1000
    dy = haversine_dist(bndg_box[1][0], bndg_box[0]
                        [1], bndg_box[1][0], bndg_box[1][1]) * 1000

    return dx*dy


def haversine_dist(lat1, lon1, lat2, lon2):
    # Haversine distance between two coordinates

    dlat = (math.pi / 180) * (lat2 - lat1)
    dlon = (math.pi / 180) * (lon2 - lon1)

    lat1 = (math.pi / 180) * (lat1)
    lat2 = (math.pi / 180) * (lat2)

    a = math.pow(math.sin(dlat / 2), 2) + \
        math.pow(math.sin(dlon / 2), 2) * math.cos(lat1) * math.cos(
            lat2)
    rad = 6371
    c = 2 * math.asin(math.sqrt(a))
    return rad * c


def nni_value(dobs, a, n):
    # Calculate NNI value:
    # I-----------I---------I
    # 0.00        1.0       2.15
    # clustered  random  regular
    return dobs / (0.5 * math.sqrt(a / n))
```

```
def mean_observed_distance(coordsarr, kdtree):
    distances = []
    for coords in coordsarr:
        d, i = kdtree.query((coords[0], coords[1]), workers=-1, k=[2])
        neighbor = coordsarr[i]
        distances.append(haversine_dist(
            coords[0], coords[1], neighbor[0][0], neighbor[0][1]) *
                1000)
        # print(coords[0], coords[1], neighbor[0], neighbor[1])
    return mean(distances)


if __name__ == "__main__":

    parser = argparse.ArgumentParser(
        description='Calculate the Nearest Neighbor Index of a dataset
            ')

    parser.add_argument('dataset', type=str,
                        help='absolute path to the dataset (must be a .
                            csv in form lat,lon)')

    parser.add_argument(
        '-b', '--bounding_box', help='choose an area bounded by 2
            coordinates in comma-separated form: lat1,lon1,lat2,lon2. If
             no value is supplied the entire area is chosen', type=
            lambda s: [float(item) for item in s.split(',')])

    parser.add_argument(
        '-j', '--json', help='export bounded area to a json file JSON')

    parser.add_argument('-v', '--verbose', dest='verbose', action='
        store_true', help='verbose output (show number of points etc)')

    args = parser.parse_args()

    #  DF must be in format lat, lon (comma separated)
    df = pd.read_csv(args.dataset)

    #  We get rid of any rows with non-numeric values
    df['lat'] = pd.to_numeric(df['lat'], errors='coerce')
    df['lon'] = pd.to_numeric(df['lon'], errors='coerce')
    df = df.dropna()

    #  Bounding box is arguments, unless not supplied
    if args.bounding_box is not None:
```

```
    boundingbox = [[args.bounding_box[0], args.bounding_box[1]], [
        args.bounding_box[2], args.bounding_box[3]]]
else:
    boundingbox = [[df['lat'].min(), df['lon'].min()], [df['lat'].
        max(), df['lon'].max()]]
bounded = box_select_coordinates(df, boundingbox)

#
#   We must put the coordinates in form [[lat, lon], [lat, lon]...]
#
npbounded = []
for i in range(len(bounded)):
    npbounded.append([bounded.iloc[i]['lat'], bounded.iloc[i]['lon
        ']])

npbounded = np.array(npbounded)

#
# We need to construct a k-n tree, as multiple queries will be made
#

print("Constructing k-d tree...")
nni_querytree = KDTree(npbounded)

#
#   We must then calculate the NNI:
#
#   Rn = D(Obs)/          where: a is the area observed, n is the
    number of points, D(Obs) is the mean observed distance to the
    nearest neighbor
#      0.5(sqrt(a/n))
#
#   Loop through all coordinates, finding their nearest neighbor
    and calculating the distance using the haversine formula
#

nni = nni_value(mean_observed_distance(
    npbounded, nni_querytree), box_area(boundingbox), len(npbounded
        ))

print("Calculated NNI value for", args.dataset, "=", nni)
if args.verbose is True:
    print("n: ", len(npbounded))



#
#   Export JSON is user has chosen to
#
```

```
if args.json is not None:
    # create JSON file
    json_file = bounded.to_json(orient='records')

    # export JSON file
    with open(args.json, 'w') as f:
        f.write(json_file)
```