Visualization Group 44 Final Report

Coen Smeets* — Mechanical Engineering — 1615947

Armando Cerullo† — Mechanical Engineering — 1554085

Jakob Haud‡ — Mechanical Engineering — 1603469

Rodrigo de Miguel § — Data Science — 1571664

1 Abstract

This paper proposes a visualisation tool that was created by following Tamara Munzner nested model [18]. Datasets for restaurants and Airbnb in New York were gathered, analysed and cleaned. After the data was abstracted, multiple tasks for the visualisation were identified. A brainstorm of possible layouts of the InfoVis tool was discussed until a final product was created. Even though the overall work was based on other scientific InfoVis studies, see Sect. 3, the visualisation tool brings some original contributions. The visualisation tool is the first of its kind in creating compact multiple views where quantitative and categorical attributes can be filtered to achieve the users' goals. This study serves as an initial investigation on how different multiple views along with filtering and multivariate idioms can interconnect into a single comprehensive tool.

2 Introduction

Link to the report: `https://www.overleaf.com/read/fmjnhjwzmxdg` New York City is filled with amazing restaurants [20] which attract many *food travellers*, people who travel and plan around food. [13] Looking for AirBnBs near hotspots for food is a hassle as it requires multitasking for both searching for AirBnBs and restaurants. The visualisation tool will make this much easier by combining hotspots for food and AirBnBs into one comprehensive tool. Datasets for AirBnBs in New York [4] and restaurant data during inspections [21] will be combined to achieve the goal. Additionally, AirBnBs in New York City are in the news as a large portion will be illegal in the upcoming future which should be added as a warning to the users. [17]
The goal of the visualisation tool is to *present* the options in a comprehensive way so that users can *explore* them and make the best choice.

The combinatory use of spatial data (location of listings and restaurants) and sequential quantitative data (listing price, service fee, restaurant scores etc.) will be crucial to achieving the goal in hand. Spatial data can as an index between the semantics and vice versa, allowing the user to alter its options and see the results instantly in a comprehensive way. This makes the visualisation tool insightful, yet relatively simple for the target user.

As the possibilities are immense and geographical location is essential for the choice of an AirBnB, visualisation is needed to present the data in a way that is understandable and easily accessible. The data of the restaurants in New York will help locate hotspots of good food from which an AirBnB can be chosen within that area.

*e-mail: c.r.a.smeets@student.tue.nl
†e-mail: a.cerullo@student.tue.nl
‡e-mail: j.haud@student.tue.nl
§e-mail: r.de.miguel.satue@student.tue.nl

3 Related Work

The two data sets in question, Airbnbs and restaurants in New York City, are both commonly used for data exploration purposes. Thus, many analyses and visualisation tools have been created already. For example, for the Airbnb data, a tool exists that *presents* the data to *enjoy trends*; the tool visualizes the Airbnb data over the years. [9] For the restaurant data set, there is a visualization experiment that won Kantar's visualization his award; [6] Furthermore, there exists preprocessing examples of the same datasets used, that will aid the process and expands the time left for visualizing the data. [5, 8, 16]. Nonetheless, restaurant and Airbnb data have always been separated which prohibits overview and comprehensiveness.

The visualisation tool will base its interactions on [15]. Multivariate idioms will be instead based on [10]. The idea is to put to practice the geocomputation that is made in the paper by combining maps idioms with PCPs. Structuring multiple views was instead based on a study by Michelle Q. Wang Baldonado, Allison Woodruff, Allan Kuchinsky [22]. Using a map as an idiom was an inspiration taken from a project by [14] Hang Yi.

4 Data Analysis (What)

As previously stated, this tool enables a user to interactively visualize New York City AirBnB listings and restaurant locations. This section documents the reasoning behind choosing the selected data as well as the steps taken to process it before implementing it within the tool.

Domain Data Specification

The Airbnb dataset from [4] was provided by the course, while the restaurant data was gathered from the results of the NYC Department of Health Restaurant inspection from 2017 [21]. The data set description can be seen in Table 1:

Table 1: Overview of the Dataset of Airbnbs

Attribute type	Attributes	Ordering
Categorical	NAME, host_identity_verified, host name, neighbourhood group, country, country code, instant_bookable, room type, house_rules, license, ID, Host ID	None
Ordinal	cancellation_policy	”flexible” to ”strict”
Quantitative	lat, long, Construction year, price, service fee, minimum nights, number of reviews, last review, reviews per month, review rate number, calculated host listings count, availability 365	sequential

Another attribute was derived and it is called ”legality”. It warns the user about a possible breach of rules when a room or apartment is rented for longer than 30 days. This was based on NYC Airbnb rules [3]. This categorical attribute was used in the visualisation tool to give a warning of possible rule breach to the user. A more detailed view can be seen in Fig. 11 Several steps were taken to pre-process the restaurant dataset, including feature selection, dropping missing values, deleting duplicates for dataframe keys, and encoding the location data, originally stored as latitude and longitude values, as geobuf data for faster manipulation [2]. The threshold applied for the most salient attributes can be seen in Table 2:

Table 2: Threshold values for outliers

Attribute	Threshold
Airbnb	
Minimum Nights	Removed for above 90 [1]
Construction Year	Removed for below 2002
Restaurant	
Grade	Removed if equal to Z, P, Not yet graded.
Score	Removed if below 0

Geobuf encoding is especially useful for plotting points efficiently [2]. The additional restaurant dataset, originally containing 20 attributes was supplemented with latitude and longitude. The 400000 addresses were used with the Openstreetmap API to convert them over the course of a week [19]. Additional attributes such as reviews, room type, price and cancellation policy were kept as these could be relevant for the user during visualisation to browse for Airbnbs that suit their needs.

Data Abstraction

The following data abstraction was performed according to the guidelines stated in Chapter 2 of Tamara Munzner's "Visualization Analysis and Design" [18]. The data consists of two static flat tables, with Airbnb listings and restaurants as items in their respective datasets, followed by their attributes. The attributes are of several types: there are categorical attributes such as cancellation policy or room type, quantitative ordered attributes such as rental price or service fee, and sequential ordered attributes such as review rate number, for which values range from 1 to 5. Moreover, geometry data is in the form of latitudes and longitudes both for the AirBnB and restaurant datasets.

5 Task Analysis (Why)

This section will be split in half. Firstly, possible visualisation tasks that the user might undertake will be analysed. These tasks will still be domain-specific and have to align with the goal of the visualisation. Finally, these tasks will be generalised into a domain-independent abstract form.

Domain Specific Tasks

The visual tool has many different uses (Sect. 7), hence the user can approach it from contrasting sides, e.g. looking at restaurants near Airbnbs or vice versa. Due to it, the goal of the visualisation is quite broad and has to be able to solve multiple different tasks that the user might encounter when using the visual tool. These tasks can vary from high-level tasks, which are connected to analysis, or low-level tasks which pertain mainly to querying. All of these tasks need to be included.

After analysing the potential tasks the users would be interested in doing and the visualisation tool would solve, a set of tasks with corresponding questions were formulated. The tasks take the whole of the data into account and use it to its full potential, encompassing its different attributes and a complementary data set. The tasks range in complexity, some take into account multiple attributes and relations between them.

- Discover areas in New-York with many restaurants (Where are the hot spots on restaurants?)
- Find the best AirBnBs in the area (What are the best AirBnBs in the area to stay in?)
- Do not choose an AirBnB which is illegal. (Is an AirBnB following the New York City laws?) [17]
- Get the trustworthiness of a listing (How trustworthy is the listing, looking at the last review, good reviews, and amount of reviews?)
- Filter out restaurants/AirBnBs on price and other factors (How do the factors influence the aspects of a restaurant/AirBnB?)
- Enables to examine what contributes to a highly rated restaurant/AirBnB (Which attributes contribute to a highly rated restaurant/AirBnB?)

Task Abstraction

The tasks could now be abstracted into a domain-independent form to make the use of visualisation principles for the visual and interaction design more straightforward. There are multiple different abstractions possible however, one will be chosen and further explained for each domain-specific task. The ordering is coherent with the ordering for the domain-specific tasks.

Discover extremes: Hot-spots are the most densely populated areas, which corresponds to extremes in abstracted terms. The extremes have to be discovered for further selection.

Browse features: The best Airbnb depends on all data for each listing. The features in terms of price, score, and location are all important. The location of these attributes is already known. Which listing, the target, has the best is unknown thus, browse is the abstraction.

Annotate dependencies: Whether an Airbnb is illegal depends on the dependency on the type of place and minimum stay length. The minimum length has to be below 1 month and the type of place has to be a whole apartment/house. [17] The one attribute depends on the other. Additionally, the tool should showcase the possibly illegal listings and the user should not have to look for them. Thus, annotate is the task abstraction at hand.

Present correlation: For example, a 5-star reviewed place can only be considered reliable when it is based on numerous and recent reviews. Thus, the correlation between attributes are important. The data should be presented so the user can make their own conclusions on the presented correlations as trust differs per person.

Discover distributions: The filtering of attributes can only be done when the distribution is known to the user. An example is an average review. Someone can demand a 5-star average but only a few listings will remain. Having the distribution to balance the number of items and the desired output is necessary.

Discover features: The tool should showcase the features a restaurant or AirBnB has and must enable to compare attributes, such that the user could understand what features contribute to their search goal (in this case highest rated restaurants/AirBnBs)

The tasks range from high to low level, which taken together will contribute to the end goal of the visual tool. High level tasks of a visualisation enable to consume or produce information, potentially generating or verifying new hypotheses [18]. These types of tasks are discovering distributions and features. High level tasks contain low level tasks, i.e. the browsing of features contributes to making discoveries about said features. Taking these abstract tasks into account, design solutions can be developed together with viable justifications for the decisions made.

6 Current Solution (How)

A and B: Brainstorming and Improving

All figures that are shown in the following section can be found in Sect. 9

First Concept

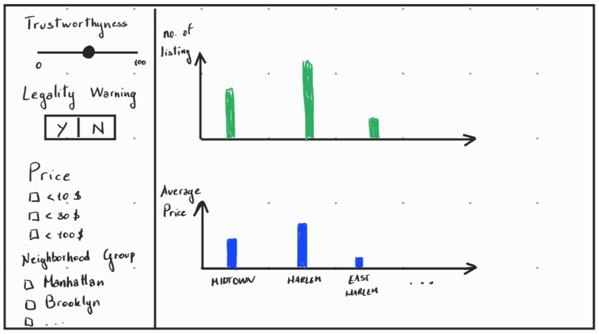

Figure 1: First Concept

The idea of Fig. 1 and Fig. 12 is to have multiple plots on the central view that show distribution and trends. To this extent the main idioms used includes bar charts and scatter plot.

Table 3: Evaluation table for the first concept.

Pros	Cons
Output panel allows the user to apply multiple filters at the same time.	Idioms is the wrong choice for spatial data. Maps would work better.
Distribution and bar chart as idiom give a good overview	Distribution analysis can be aided with basic summary statistics
Simple idioms for categorical data such as neighbourhood make the plot simplistic and easily understandable	Output panel might be too crowded. Gestalt's principle should be used to create areas to group options (e.g. price and neighbourhood) [18].
The visualisation is versatile. More plots could be added, axes could be changed, different idioms could be shown	Colour channel is superfluous in the bar charts.

In conclusion, the plot made use of idioms such as scatter plots, bar charts in the central view but it resulted in an overcrowded plot with no use of spatial data. Furthermore, the idioms chosen mainly enable to solve low and mid-level tasks. High-level tasks such as making discoveries about the whole data not only a single or a few attributes is not possible.

Second Concept

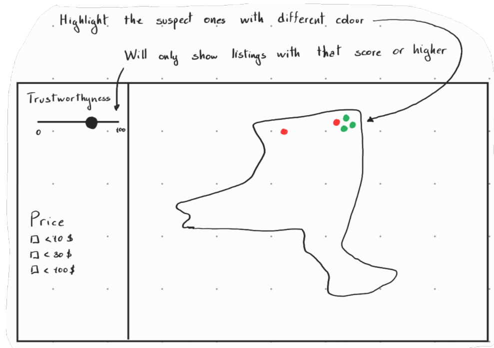

Figure 2: Second Concept zoomed out.

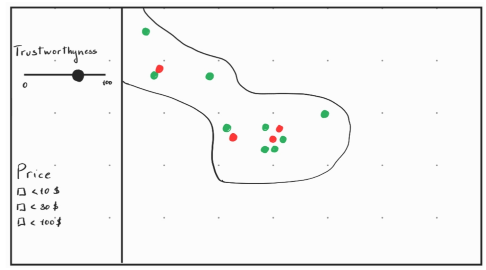

Figure 3: First Concept zoomed in.

Fig. 3 and Fig. 13 makes better use of spatial data by using a map to encode the latitude and longitude of the listings. Consequently making exploration and discovery tasks more straight forward. The idea is to have colour as channel to encode the legality of the listings. Red is used when a listing is likely illegal.

Table 4: Evaluation table for the second concept.

Pros	Cons
Using position as a mark for spatial data eases reading the graph for the users.	More channels could be used to encode attributes of the output panel in the same plot.
Allowing the user to zoom in the map make the plot more interactive and allows the user to find more geographically accurate information	The plot does not show distributions of any of the attributes and not even summary statistics.
Colour channel is used appropriately to encode data (red = possibly illegal, green = likely legal)	The plot could encode trustworthiness in the central view.

As a direct improvement to Fig. 1 the output panel is less occluded since most of the data moved to the central view of the plot. The data ink ratio was taken into account by making the map not show any geographical data and being fully blank. In addition, semantic zooming allows the dots to scale as the user zooms, eliminating possible occlusion. In conclusion, the plot is starting to shape up. However further simplification should be tried. Using a map as an idiom enables the user to perform higher level tasks, since changes on the data can easily be visualised.

Third Concept

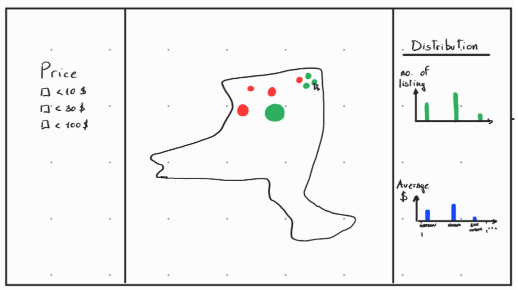

Figure 4: Third Concept main visual.

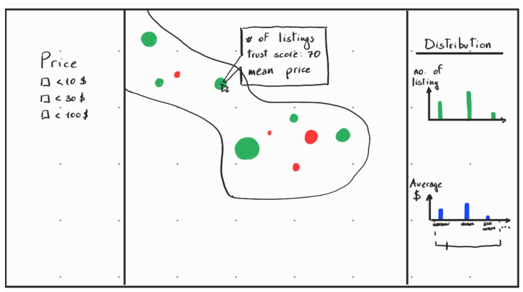

Figure 5: Third concept after the user zoomed in.

The idea of Fig. 4 (Fig. 15) is similar to Fig. 3 but trustworthiness is encoded in the central view using the area channel. The level of tasks performed by the visualisation stays the same.

Table 5: Evaluation table for the third concept.

Pros	**Cons**
Showing listing information by hovering the pointer is a great alternative to decrease users' cognitive loads [18].	The interference of the area and colour channel should be removed.
Adding more bar charts showing distribution helps the users eye-ball summary statistics (by correctly applying Tufte's principles).	A slider might be better for the price instead of tick boxes.
Information while hovering is a good alternative to showing the entire history	Colour used superfluously in bar charts.

In conclusion, Fig. 4 shows separability as color and area (size) interfere.

Chosen Concept

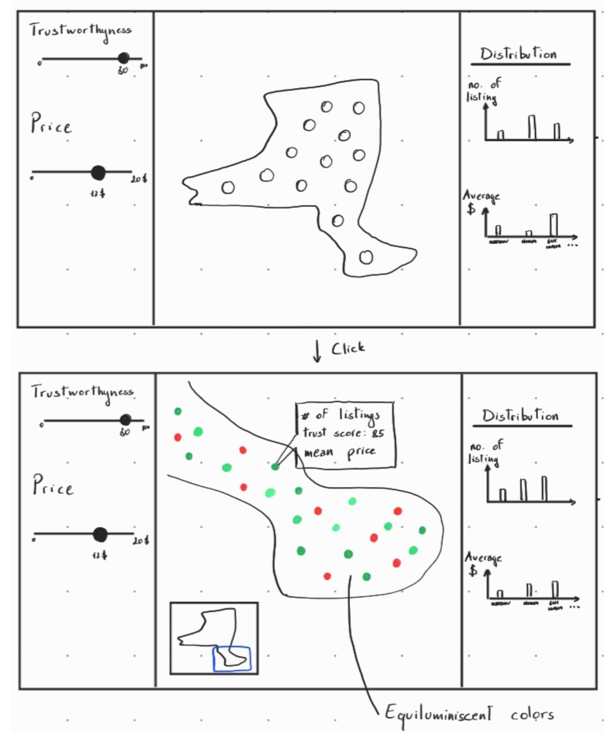

Figure 6: Final Design zoomed out

The chosen concept (Fig. 17) will implement all that was learnt in the brainstorm phase. In this case a map from Fig. 3 and Fig. 4 was deemed best to illustrate the spatial data. The user in the default view can select a threshold for price and trustworthiness and on the righter most part, distributions for listings and price according to neighbourhood can be seen. After a cluster is clicked, the map semantically zooms in so that the user can see listings. The colour use is the same as in concept two. Since semantic zoom is applied, a mini-map was added to make the user aware of its position. Hovering with the mouse will give additional information on the selected listing. The choices made can be justified using the principles as defined in Munzer's Visualization. Analysis & Design. [18]

Tufte's principle: The data ink ratio has been optimised for the bar charts and map compared to the aforementioned concepts. Additionally, the bar charts have correct vertical scaling avoiding possible bias.

Gestalt's principle: The visualisation is divided in three distinct sections making use of the proximity principle. The visualisation also uses two main colours which makes the user perceive them as two distinct groups as intended (Similarity principle).

Colour: The colours were made equiluminescent. The hue will outline the attribute 'trustworthiness' since it is sequential data.

Semantic zooming: Semantic viewing will be aided by a general overview of the zoomed-in detailed graph to the overall city of New York.

In terms of idioms the exploration of the concepts allowed to achieve the conclusions in Table 6

Table 6: Idioms use

Goal	Used Idiom	Justification
Show geographical positions	Map with clusters	Spatial data (lat and lon) can be easily be conveyed in a map.
Filtering quantitative data	histograms and violin plots	histograms give a direct visual of the distribution and violin plots show its summary statistics.
Filtering categorical data	Check boxes	Distribution and summary statistics are not possible for this kind of data. The effects of filtering can be and should be directly visible in the map.
Multivariate Idioms	PCP and density plot	A PCP can more easily filter all the quantitative data in the dataset.

These idioms individually enable performing low level tasks like discovering extremes, browsing features, annotating dependencies and presenting correlations. Combining them enables the user to discover distributions and features of the data, the key idiom that binds everything together is the map, which makes visualisation clear and the results/outcomes from the use of filters can be seen instantly. Using a map as an idiom was taken from [14] and is mentioned in Sect. 3. Nonetheless, it was thought that its interactivity and use could be extended. For this reason, it was chosen to make the map interactive through filtering and manipulations (semantic zooming, panning, constraining, and selection). Some aspects of the chosen concept can be further highlighted in their inherent visual principalities. With expressiveness, the two distinct colours help the positions to pop out even more. The separability has also been taken into account by not using the area as a channel combined with colour. Making the plot interactive by adding the zoom manipulation, the position accuracy becomes more precise and data can be filtered by an intuitive method, zooming on a map. Effectiveness is central in the choices; Position, expressed in latitude and longitude, has the channel of position which is ranked as the most effective.

Final Product

The most important packages used to do build the visualisation can be seen in Sect. 6.

- **Dash**: Creates the framework for the rest of the visualisations. Provides an interactive local website.
- **Plotly**: Creates basic interactive graphs but has limited flexibility and has built-in functions for Dash specifically.
- **Dash-leaflet**: A light wrapper around React-Leaflet. [11] Creates interactive spatial data graphs.
- **dash-bootstrap-components**: A library of bootstrap studio components for Dash. [12]
- **Dash-loading-spinners**: a library that extends the functionalities and looks of the dash loading component. [7] Used for changes that fall outside of the the categories of *perceptual change* and *immediate response*.

The final products consists of multiple interactive views. A general schematic describing it can be seen in Fig. 7 and Fig. 21:

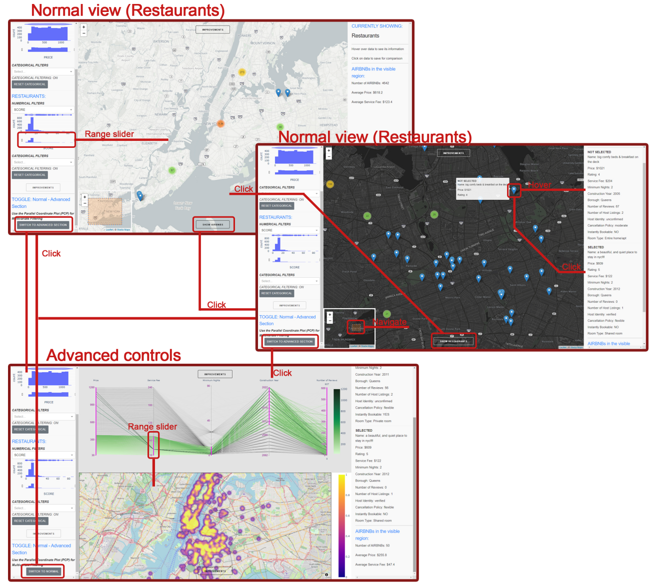

Figure 7: Schematic of the tool.

In general, the layout was based on Proximity as Gestalt Principle [18]. The controls are all grouped together on the left and the selection information is all on the right side of the screen. Each of these views will be treated more in-depth in the remainder of the chapter.

First Multiple View

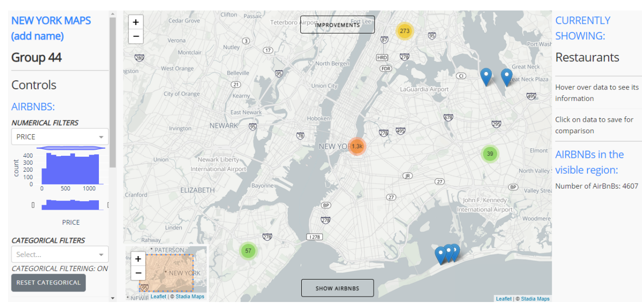

Figure 8: First Multiple View.

This first view in Fig. 8 and Fig. 18 makes use of a map, histograms, and violin plots. The histogram and violin plots are interactive and used for filtering. The results of such filtering can then be seen on the map. Filtering principles such as showing the distribution of filtering via violin plot were imported from [15] and [18]. For the map, possible manipulations include semantic zooming and select. The user is able to zoom in on the listings and select them as suggested in [15]. Once a listing is selected, its information is reported on the right side of the screen. This allows for comparison with other listings. The reason for this is to decrease the cognitive load of the user as seen in [18]. Other general statistics such as the number of listing in an area, the average price and service fee in the area are provided so that the user has real-time information on his interactions.

If filtering values are changed, a loading icon appears on the right side of the screen to notify the user that the process takes longer than the response time they expect [18].

Second Multiple View

The view can be switched to see Airbnb listings by pressing the button Show Aibnbs leading to Fig. 9 and Fig. 19.

Figure 9: Second Multiple View for Airbnb listings.

The colour of the map turns to black creating a high contrast from the previous view. This is based on the expressiveness principle [18]. The dark color reminds the users of nighttime and therefore a place (Airbnb listing) for sleeping. The light colour of the restaurant matches the food trip that is ongoing during the day during light hours. This also follows the rule of "Self-Evidence" [22].

Advanced Multiple View

Figure 10: Second Multiple View.

The advanced view in Fig. 10 and Fig. 20 consists of a PCP coordinated with a density plot. The PCP is the filtering device and the results can be seen in the density plot. When the view is switched to advanced, the map is replaced by the PCP and density plot following the multiple views principle of "Attention Management" [22]. More information can be extracted off of the density plot as showed in Fig. 11 and Fig. 22:

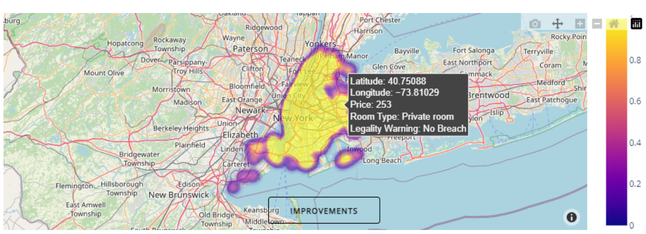

Figure 11: Hovering on the density plot.

This is an extension to combing spatial data with PCP as studied in [10]. Encoding the data in other idioms was also suggested in [15]. An overview of the main design choices made for the advanced view can be seen on Table 7.

Table 7: Design choices for the Advanced View

Category	Design Choice	Explanation
PCP	Transparency and Interaction	The axes of the PCP can be moved and the plot is interactive as a filtering range can be input. Opacity was applied to aid readability
Color Scale	Color choice	The PCP color bar was chosen sequential and with different shades of green as humans are most sensitive to it.
Density Map	Color choice and filtering	By using another idiom, the "out of sight out of mind" effect due to filtering is eliminated. Salient data can be gathered by hovering
Density Map	Geometric zooming	Geometric zooming is used to help the user identify listings when zoomed in and see what the filtering changes in the big picture.

Performance

Having two datasets, with the ability to switch between maps while performing the necessary tasks, created extra complexity in the implementation of the visual tool. While these were mostly overcome, some decisions had to be made between prioritizing the end goal of the interface with the overall layout and its features. Individually, i.e. restaurants and AirBnBs can be analyzed in a way that conforms to the visualisation principles in [18]. The implementation of two separate maps made it hard to create a combined way of visualizing selected AirBnBs and restaurants, to directly compare their location and features in a visually convenient way. Nevertheless, by switching between maps, the main goal of the visualisation tool, of having a comprehensive way to compare and explore options is certainly possible.

The filtering process, which is done individually in a loop using geojson data, is relatively fast and takes approximately half a second. However, the major bottleneck is the performance of certain libraries, such as the plotly library for parallel coordinates plots (PCP), which is not optimized for large data and lacks options for opacity and clustering. Additionally, the interconnectivity of the app results in a large callback with multiple inputs and outputs, which can lead to multiple firings and added delays in the overall performance, since the app cannot be used when the callbacks have not yet finished. A loading screen was added, to notify the user about the processing and make the wait more bearable for them [18].

All improvement points were also visualised as 'improvement buttons' in the visual interface. They are not part of the final product, but rather clarify the points that could be worked on further in the future. Each button goes more in depth into what could be implemented further.

7 Use Cases

The visualisation tool was made keeping the tasks and goals of the user in mind. Consequently, the use cases of the visual interface cover the defined tasks, ranging from analysis cases to simple queries. The user has great freedom when using the interface, they can approach their problems from different angles, making the interface versatile. The most important use case is the one that is related to the user's domain:

Does the tool effectively help people look for AirBnbs near restaurant hotspots? The visualisation tool manages to achieve the goal through three coordinated multiple viewsSect. 6. This combined with filtering and multiple manipulations in the map (selection, panning, constraining) allows the user to choose the

listing according to their wants and needs.
Other smaller USE cases within the main goal include:

Airbnb

Where are AirBnBs most present? This use case could be solved by the visualisation tool. The highest concentrations of Airbnb are in the Manhattan, Brooklyn neighbourhood group. The use of a map with spatial data and filtering makes the use case easily achievable.

What is the distribution of quantitative data? Continuous distribution for price, service fee, and construction year: This use case can be achieved by the visualisation tool. Filtering of quantitative data is in the form of histograms and violin plots showing summary statistics and general distribution shape. The listings for the Airbns show a surprising continuous distribution for their price, service fee, and construction year. Minimum nights, number of host listings, number of reviews show a right-skewed distribution.

Where are listings cheaper? Cheap listings can be located by filtering prices and simply looking at the map. Furthermore, on the right hand of the screen, the total number of Airbnb is report. Potential clusters include the South Williamsburg area and Manhattan lowers east side.

How many listings have their Hosts' ID verified? The visualisation tool can give an exact number by filtering and reporting the number on the right-hand side of the screen. Nearly half of hosts' identities are not verified. As a bonus, if a listing is selected, this information is reported to the user in the right hand side of the screen.

How many legality warnings are there? Nearly every home/apartment listing is possibly breaching the NYC rule stay of 30 days [3]. This was discovered by using geocomputation from Sect. 3 [10]. The visualisation does not completely fulfill the use case as it cannot give an exact number but can give the warning per listing. This is shown in Fig. 11.

Where are the old and new listings located? The visualisation tool can achieve this in the second multiple view via the controls on the left hand side. Both the newest and oldest listings are all located in the Manhattan and Brooklyn district.

What contributes to making an AirBnB listing expensive(above 1000 dollars for minimum nights less than 5 days)? The app can give some very limited information and does not fulfil the USE case. Filtering can partly achieve this to set the right bounds but the user has to check listing per listing to read the description.

Restaurants:

Where are restaurants mostly present? The visualisation tool can achieve this goal by using the control on the left hand side and the map. The highest concentration of restaurants is in the Manhattan and Brooklyn neighborhood groups. An exact number is not shown on the right-hand side of the tool unlike with Airbnbs.

How many restaurants have violated sanitary codes? Filtering in combination with the map can answer the question. Out of 9.6 thousand restaurants, 9.5 thousand have had sanitary violations cited against them. Exact numbers cannot be provided.

How many sanitary violations are critical? Categorical filtering and the map allow for this. Out of 9.6 thousand restaurants, circa 4.7 thousand (nearly half) have violated sanitary regulations critically. They are mostly concentrated in the Brooklyn and Manhattan neighbourhood groups.

Where are the highest-rated restaurants? What attributes contribute to a high rating? Quantitative filtering with histograms and violin plots shows that the highest-scored restaurants are located in the Queens, Manhattan, and Brooklyn neighbourhood groups. Categorical filtering and the comparison option of the tool enable comparison restaurants to evaluate what are the most important attributes that high-ranking restaurants have. For example, do these restaurants have violations, if so, are they critical?

Are there any cuisines with high grades? Italian and American restaurants nearly all have A as a grade. This can observed by filtering.

8 Conclusion and Future Work

New York City is a popular destination among tourists, particularly those who prioritize food experiences when planning their travels [13]. The search for accommodation near food centres can pose a challenge for these individuals. In order to address this issue, a visualisation tool was created. The inspiration for this tool is rooted in the lack of a comprehensive and convenient tool to aid in the search process. An iterative design process, guided by it [18], is employed to develop a user-friendly visual interface for the tool.

Multiple tasks were derived which were made into abstract form, ranging from high-level analysis to low-level querying tasks. Multivariate filtering techniques enable the user to manipulate the data to their liking and the center piece of the visual interface, the map, uses spatial data to illustrate the findings for the user in a familiar way. The user is able to see data about restaurants and AirBnBs, understand their distributions, and correlations and compare them.

The tool has many use cases Sect. 7 that are intertwined with the tasks of the user. These could be simple filtering uses, finding legality warning in AirBnBs, seeing how many hosts' IDs are verified, and filtering restaurants that have violated sanitary codes. Mid-level and high-level uses come from tying the results of the low-level tasks to the map and using multivariate filtering to make conclusions about restaurants, AirBnBs and their features in New York. For example, concluding which attributes contribute to a high rating restaurant or accommodation and understanding where they are most present.

The scope of this project did not enable the team to fully finalise the visualisation interface and fully undergo the validation process proposed by [18]. In terms of the interface, the visualisation package Plotly, sets some constraints around the complexity of the quantitative filtering system. After filtering is applied and another attribute there is a loss of coherence between the filtered data and the histograms. Overlaying restaurants on the AirBnB map and vice versa was not implemented due to this reason. Visual separation of selected items and the connection between the density plot and maps has to be improved. Furthermore, the interconnectivity of the app results in many large callbacks with multiple inputs and outputs, leading to added delays in the overall performance of the algorithm.

Validation of the visual tool was not fully completed. While analysing the algorithm by looking into and reducing computational complexity was mostly successful, task abstraction and domain-level validation were not conducted. As future work, field studies should be done to document and analyse the users' usage of the system. Furthermore, tests on target users should be conducted to measure adoption rates.

REFERENCES

[1] The 90-day airbnb rule and what it means to you [london, 2022].
[2] NPM Documentation: Geobuf.
[3] Airbnb. New york, ny.
[4] A. Azmoudeh. Airbnb Open Data, 8 2022.
[5] P. Charambira. NYC Restaurant Inspections Analysis, 4 2018.
[6] W. Chu. NYC FOODIVERSE - a data visualization experiment about restaurant in NYC, 2017.
[7] G. Down and T. Begley. GitHub - glsdown/dash-loading-spinners: Fun and funky spinners for use with Dash apps.
[8] S. Drajlin. Exploratory Data analysis of Airbnb in New York, 11 2022.
[9] C. Dwarshuis. Airbnb In New York City.
[10] R. M. Edsall. The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics Data Analysis*, 43:605–619, 2003.
[11] E. H. Eriksen. Github Dash-leaflet, 7 2022.
[12] Faculty.ai. Dash Bootstrap Components.
[13] FOODANDROAD. What Is Food Tourism? — Definition And Examples, 7 2021.
[14] Y. Hang.
[15] J. T. S. M. I. Ji Soo Yi, Youn ah Kang and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 13(6):1224–1231, 2007.
[16] N. T. C. Lai. Aribnb-preprocessing + EDA, 9 2022.
[17] C. Mench. NYC's Strict New Airbnb Regulations Could Affect Your Next Trip. 11 2022.
[18] T. Munzner. *Visualization analysis and design*. 2014.
[19] OpenStreetMap Foundation. Openstreetmap, 2022.
[20] T. Sietsema. Why New York is one of America's best food cities, 9 2015.
[21] The City of New York. NYC Restaurant Inspections, 8 2017.
[22] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '00, p. 110–119. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/345513.345271

9 APPENDIX

Page intentionally left blank

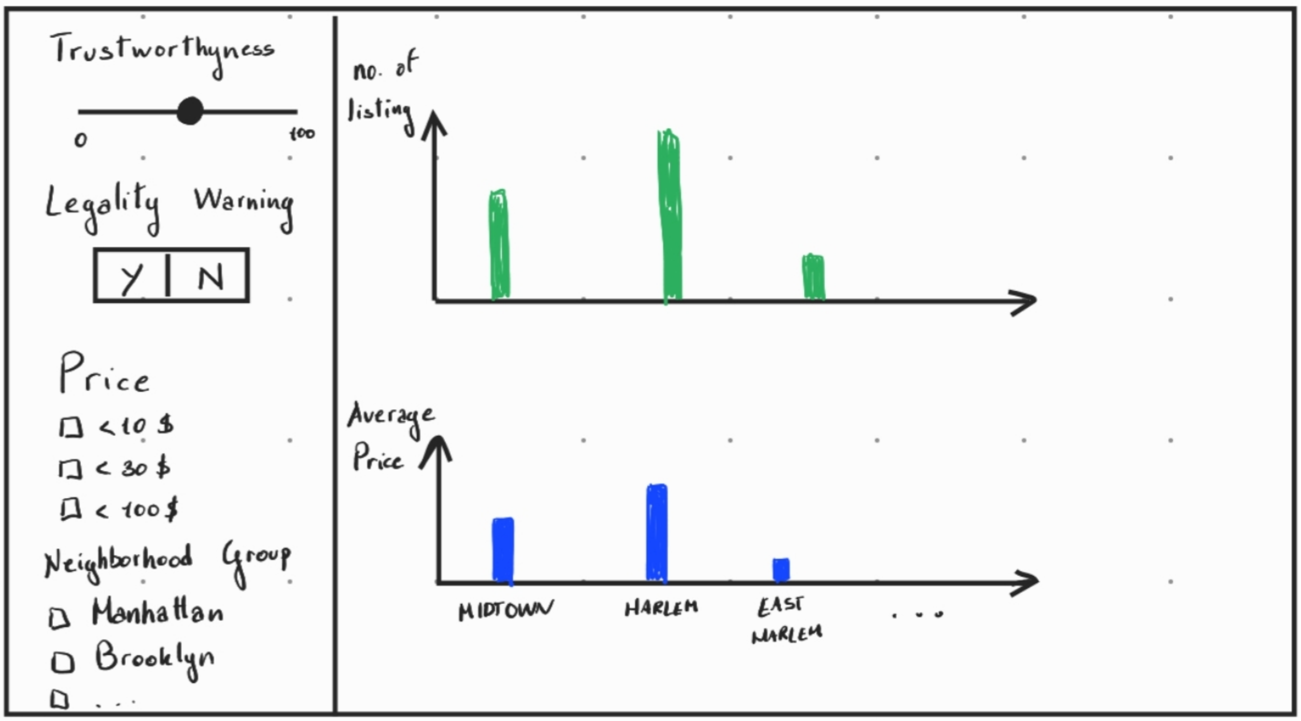

Figure 12: First Concept.

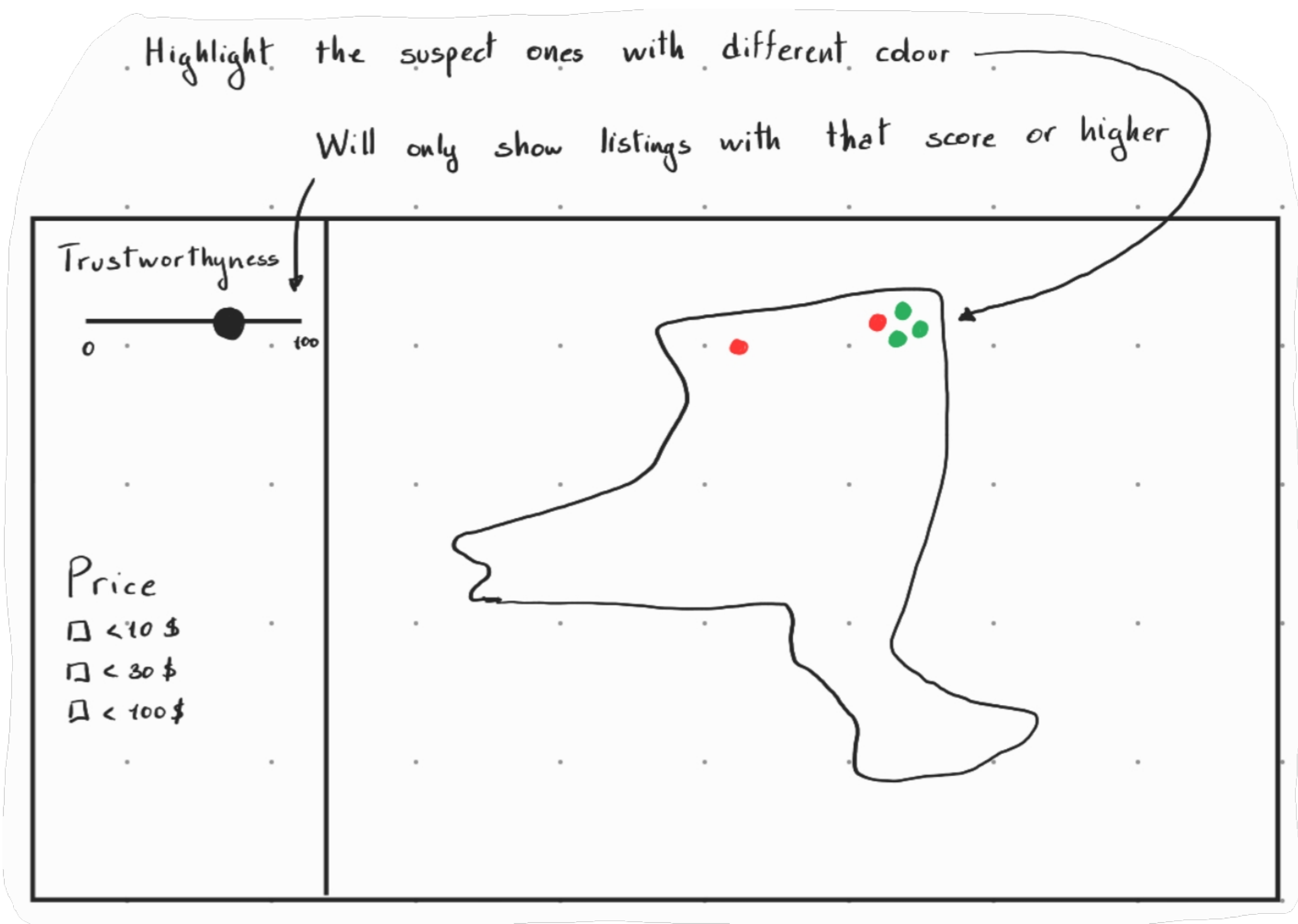

Figure 13: Second Concept.

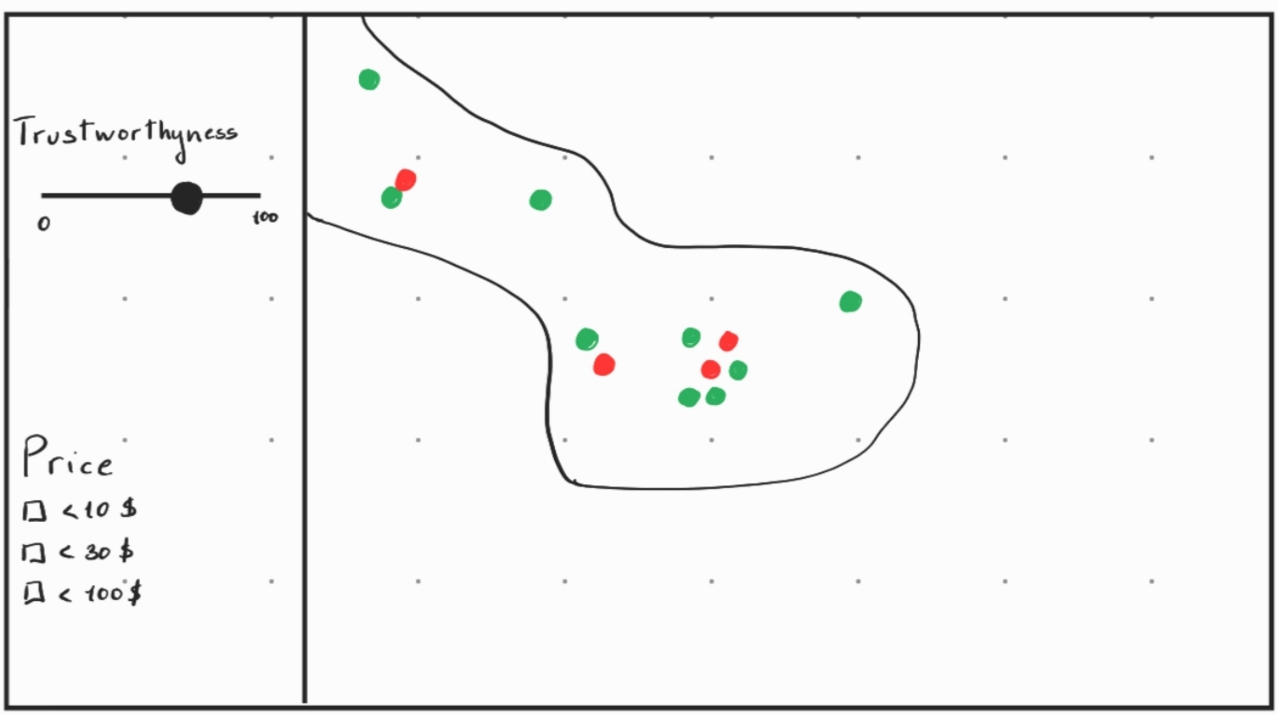

Figure 14: Second Concept Zoomed in.

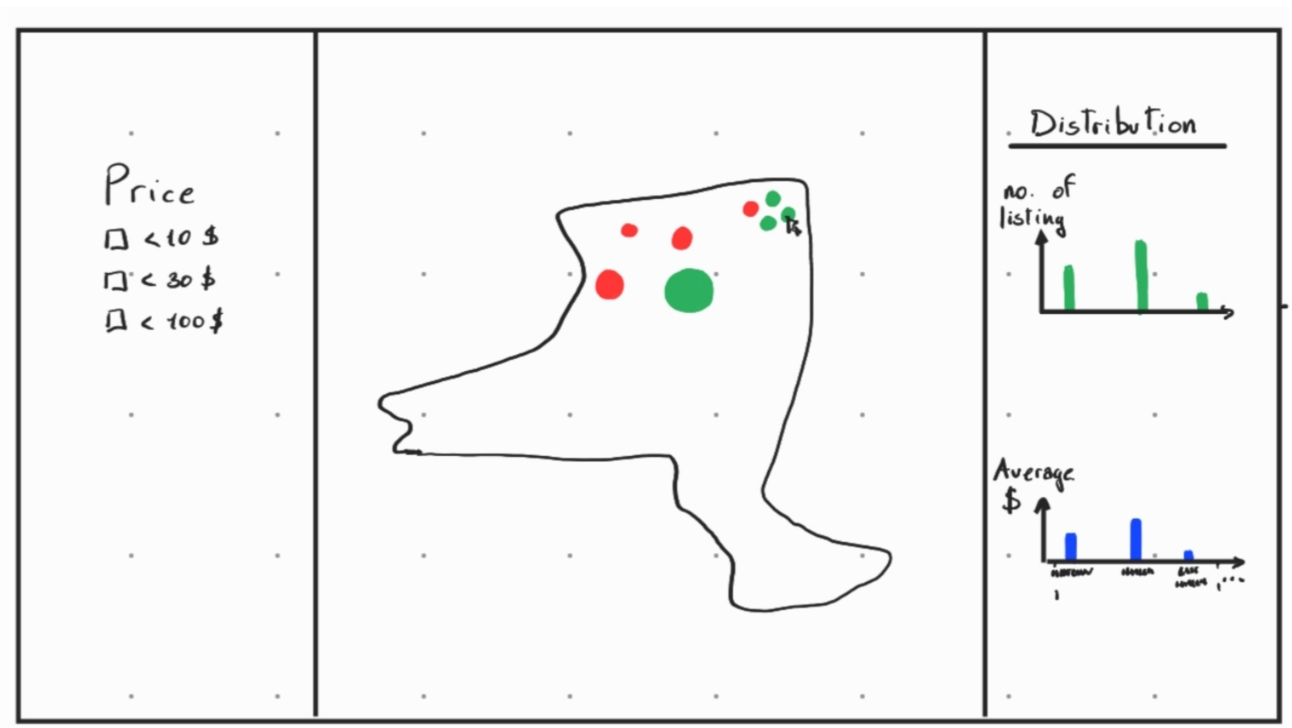

Figure 15: Third Concept.

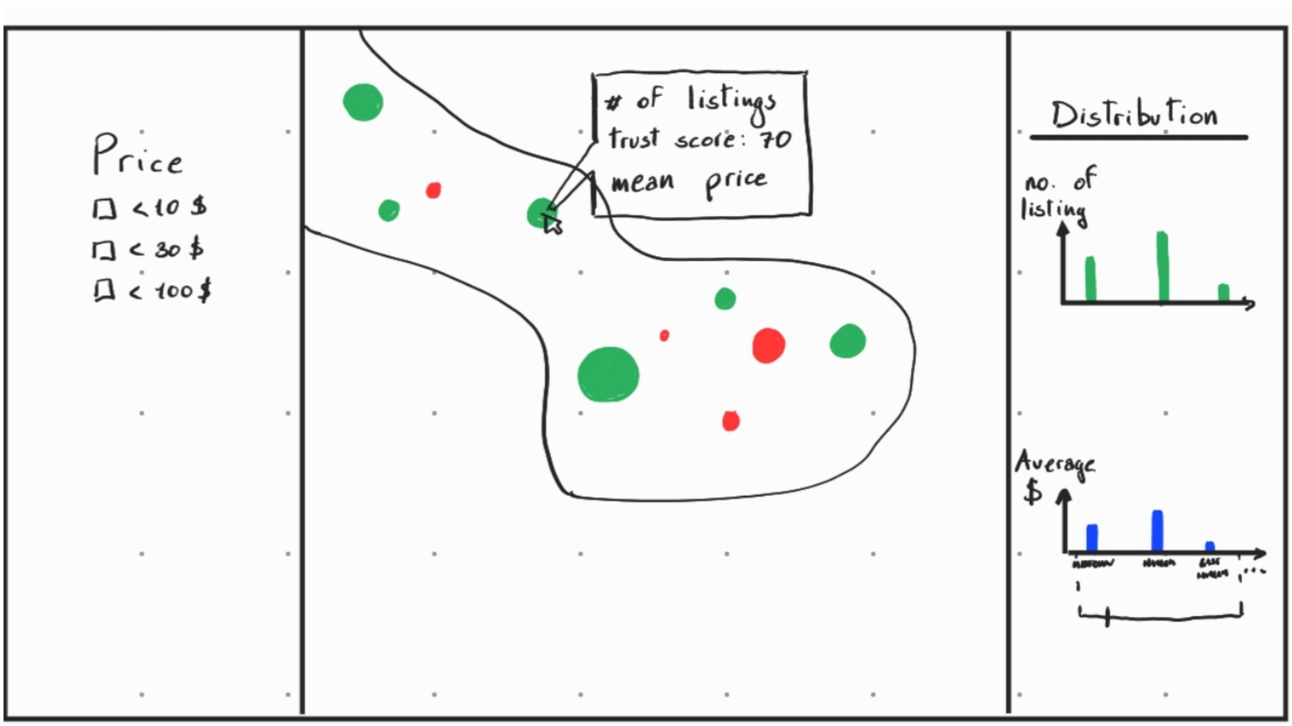

Figure 16: Third Concept Zoomed in.

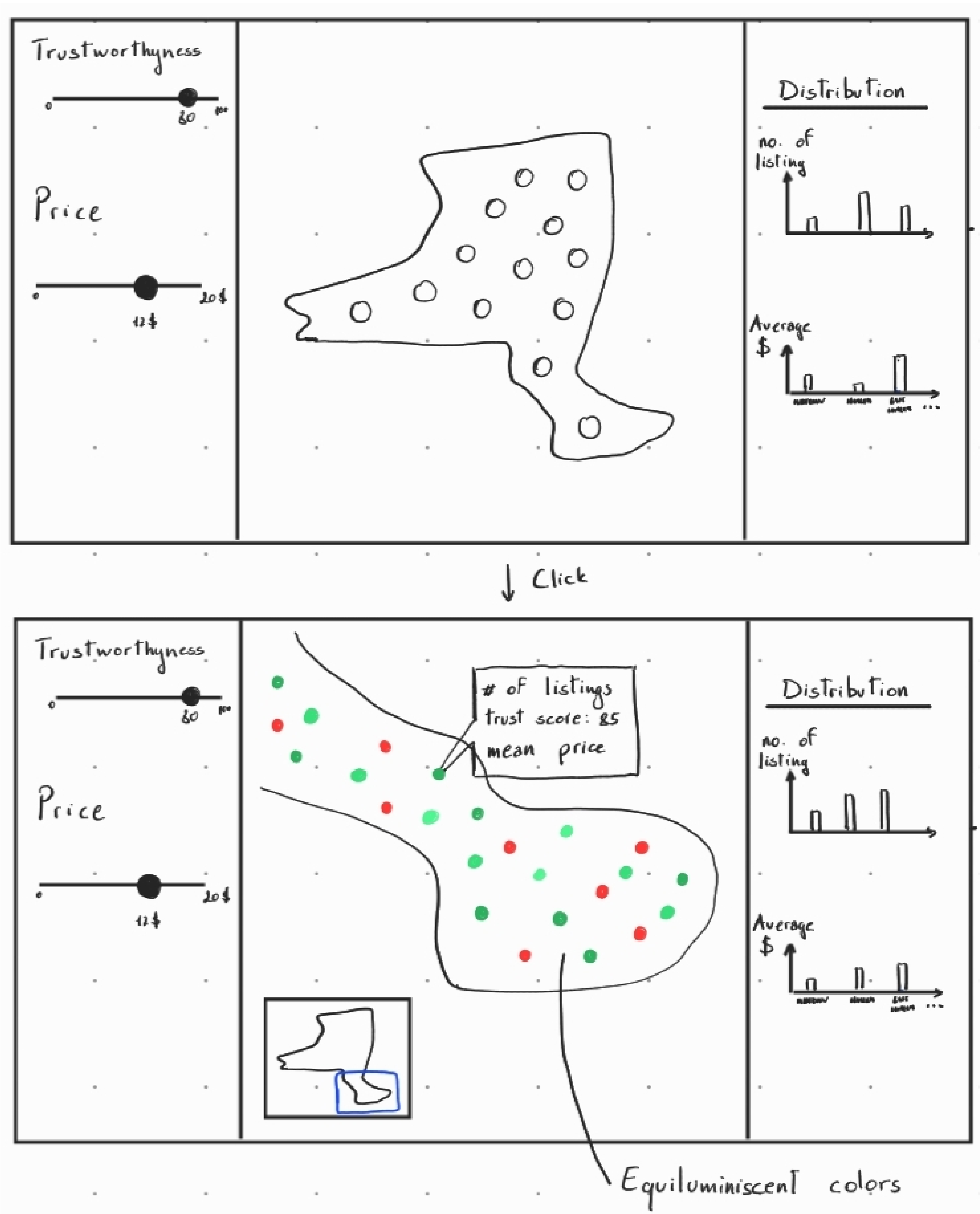

Figure 17: Final Design zoomed out

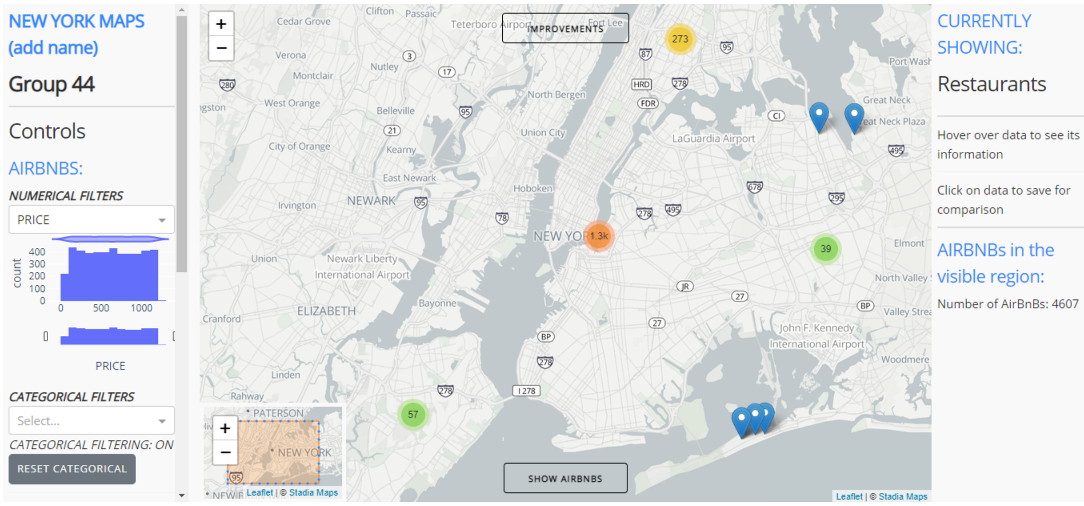

Figure 18: First multiple view

Figure 19: Second multiple view.

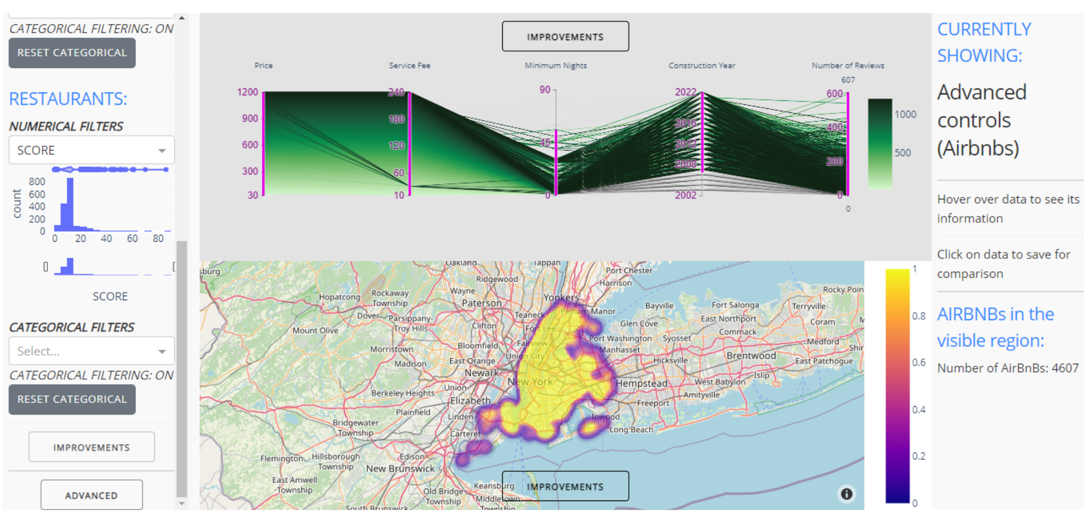

Figure 20: Advanced multiple view.

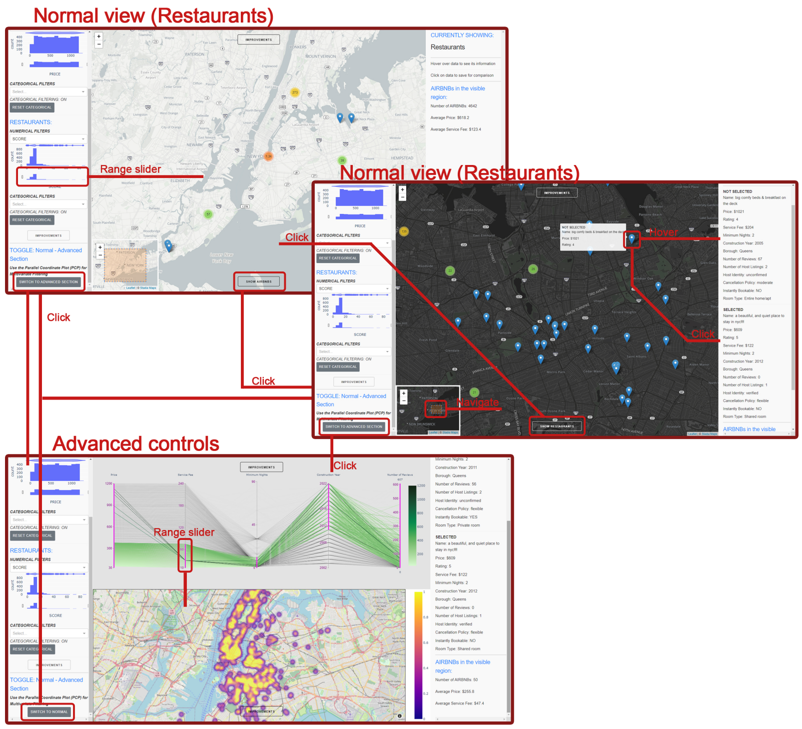

Figure 21: General layout tool.

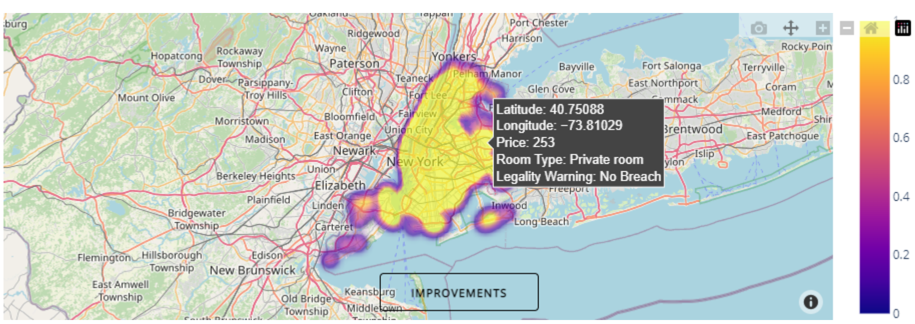

Figure 22: Hovering information.