Visual Analytics 2022/2023 - Project Report:
Italian Internet Analysis

Marco De Luca - 2017104

June 7, 2023

Contents

1 Introduction

Our society is developing to be faster everyday. Everything needs to be done better and faster, and the needs for a high-speed internet increase everyday, becoming a critical issue in many scenarios. In Italy, the situation is particularly problematic, as the country is not keeping up with the developments in other countries, which will prove to be a key issue in the years to come. This is further exacerbated by the fact that Italy's territories tend to differ quite a lot from each other, both culturally and geographically. This leads to an high variance in the availability of fast and reliable internet service in all the various zone of italy. As an end result, the so called "Digital Divide" is becoming an important issue that affects millions of people and businesses across the country.
One key event in the development of the Italian ultra-broadband network infrastructure is the institution of the Ultra-Broadband Strategic Plan (BUL) [1], approved by the Italian Government on 3 March 2015, in order to reduce the existing infrastructure and market gap. The BUL plan consists in building a publicly owned network that will be made available to all operators who want to activate services to citizens and businesses. The information related to the BUL works is freely available on their website, allowing anyone to check on the state of works in the various cities in Italy.
To help understanding this critical issue, a visualization system has been developed to help users track and understand internet speeds across Italy. In addition, the system allows to check the state of the BUL plan works in every city in the country. The final tool is an interactive visualization providing users with a graphical representation of internet speeds, enabling them to get an accurate overview of the dataset. All the visualizations are interactive and coordinated between each other. Although the visualizations are mostly intuitive and kept as simple as possible, the overall system can still result complex to a more basic, non-experienced user. Indeed, the design of the application is directed towards an experienced user such as a manager of a company looking for a zone with a good internet connection (or a zone with planned works to improve the network) where to place a new division; another possible user could be a manager of a Telecommunication company or in the Public Administration looking for insights on the zones more in need of investments on the networks.
As previously said, the final dataset contains the BUL plan informations, joined with the internet speeds data. The latter is gathered from the Ookla Open Data Initiative [5], and it will be further explained in this report. In addition, this report provides an overview of the design, development, and implementation of the visualization system, as well as its potential impact on internet speeds and accessibility in Italy.

2 Dataset

The final dataset is composed by data coming from two sources: the Ookla speedtests data for the internet speeds[4] and the BUL dataset [2] for the current (Q1 2023) state of works in Italy.

2.1 Ookla's speedtest data

Ookla is a company specialized in network intelligence and connectivity. One of the main platform they're known for is their free speedtest website, to which everyone can connect to check their own internet speed. This website is very well known and used, with hundreds of millions of speedtests taken on the Ookla platform each month. Ookla keeps track of the speedtest results, producing a dataset which is made available via their Open Data Initiative.
The raw data is aggregated into tiles, the size of each one can be estimated in meters, approximately 610.8 meters by 610.8 meters at the equator. The data is downloadable in a few formats, but for the purpose of this visualization system it was downloaded as a shapefile: this is a widely-adopted format for sharing geospatial data, supported by nearly every GIS-capable software and library.
The original dataset, updated to January 1, 2023, contains the following adjoining attributes, as described by Ookla:

Field name	Description
avg_d_kbps	The average download speed of all tests performed in the tile, represented in kilobits per second.
avg_u_kbps	The average upload speed of all tests performed in the tile, represented in kilobits per second.
avg_lat_ms	The average latency of all tests performed in the tile, represented in milliseconds.
avg_lat_down_ms	The average latency under load of all tests performed in the tile as measured during the download phase of the test. Represented in ms.
avg_lat_up_ms	The average latency under load of all tests performed in the tile as measured during the upload phase of the test. Represented in ms.
tests	The average download speed of all tests performed in the tile, represented in kilobits per second.
devices	The number of unique devices contributing tests in the tile.
quadkey	The quadkey representing the tile.

2.1.1 Preprocessing of the dataset

The dataset contains tiles from all over the world. Additionally, the tiles don't have a real geographical meaning as they are not associated to any country or city. The first operation was therefore to remove the tiles not related to Italy and then to associate each of the remaining tiles to the corresponding city. This was done using Python, and more specifically the *Geopandas, Pandas* and *numpy* libraries. The first library allowed to compute the mapping between each tile and the svg of the corresponding italian city, by using a map of Italy represented through a GeoJson file.

Then, pandas and numpy were used to clean up the dataset and remove the unnecessary attributes, namely *avg_lat_down_ms, avg_lat_up_ms, devices* and *quadkey*. In addition, the download and upload speeds were changed as to be stored in MBps rather than KBps. Lastly, in order to reduce the computation at runtime, three datasets were produced in total: the first one grouping the tiles by city level, a second one grouping at province level and a third one at region level.

2.1.2 Reliability of the Ookla dataset

The internet speedtests data from Ookla can be unreliable, mainly for two reasons, both related to the data source itself. The first reason is the fact that the dataset is only filled with the results of the speedtests performed by the users. This means that the data is not collected using a statistically reliable method, and therefore there might be issues with it: for example, some cities' speedtest data is based over a very low number of tests (in some cities even less than 10), which means their values are unrealiable.
The second issue with the dataset lies in how the speedtest itself is performed: the tests are conducted from the end user's device itself, without accounting for any possible issue internet-related: for example, an user might perform a test from a device connected via wifi to the router, while standing quite far from it; or the test might be performed while the network is under heavy load (other people downloading movies, bad weather conditions, etc.). This last issue in general tends to skew the results to be worse (lower speeds, higher latency) than they actually are. Both the issues should be considered when using the visualization system on cities with a low number of tests.
As a consequence to these issues, rebuilding this visualization system using official datasets from the internet providers in each city would greatly improve its reliability. This is especially true if the data populating the dataset was collected by executing tests directly from the modem itself, as this would prevent most of the issues previously described. Unfortunately, there are no such datasets on the internet speeds available to the public. One last observation can be made about the second issue described: while it is true that the speedtest results are worse than the real values of the end user's internet connection, it is also true that the results represent the value that the user actually experiences: in other words, they are the real values affecting the user's connectivity. If one were to obtain an official speedtest dataset from the internet providers, it could prove interesting to compare the two, in order to evaluate the possible impact of the end user's conditions.

2.2 BUL's plan data

On 3 March 2015, the Italian Government approved the Italian Strategy for Ultra-Broadband, in order to reduce the existing infrastructure and market gap, through the creation of more favourable conditions for the integrated development of fixed and mobile telecommunications infrastructures; such Strategy represents the national framework for public initiatives to support the development of ultra-broadband networks in Italy. The intervention consists in building a publicly owned network that will be made available to all operators who want to activate services to citizens and businesses.
The BUL raw data is freely available on their website (in both italian and english), and generally speaking it consists of the latest state of works for each one of the italian cities. The BUL plan also consists of developing a wireless network, but for the purpose of this system only the fiber network was considered.
The original dataset consists of the following twenty attributes and is downloadable as a .CSV file:

region_name	province_name	city_name	city_id
fiber_work_status	wireless_work_status	pcn_route	pcn_sede_id
pcn_sede_name	pcn_work_status	pcn_direttrice	pcn_ordine_direttrice
pcn_cab_transitorio	ui_piano_base	ui_piano_integrativo	fibra
fwa	piano_of_fibra	piano_of_fwa	piano_of_fibra_2021

2.2.1 Preprocessing

For the purpose of the system most of the fields in this dataset are not important. Indeed, the dataset was filtered, removing everything but the *region_ name, province_ name, city_ name* and *fiber_ work_ status* fields. The content of this last one field were in italian so they had to be translated. Finally, some cities had an empty fiber_work_status field, which was replaced with an "Unknown" status. After these operations, the dataset was joined with the cleaned up Ookla speedtest dataset.
The possible states for each city are *Unknown, In definitive planning, In executive planning, Scheduled, Being implemented, In progress, Being tested,* or *Done*.

2.3 Final dataset

Eventually, the final Ookla dataset contained 238455 tiles, each one with the 8 attributes *geometry, city, province, region, downloadSpeed_ mbps, uploadSpeed_ mbps, latency_ ms,* and *tests*. These were then grouped up by city, removing the geometry attribute (not needed anymore after the *groupBy* operation) and replacing it with the *stateOfWorks* attribute describing the BUL state of works for each city. Lastly, in order to reduce the computations at runtime, two more datasets were produced, grouping the data by province and by region. Regarding the state of Works, the grouping operation was performed by counting the occurrences of each state in every province/region. It must be noted that the downloadSpeed, uploadSpeed and Latency attributes were grouped by computing the weighted average of the attribute's value over the number of tests.
In total, the preprocessing operations took around 30 minutes. The final city dataset contains 7783 tuples and 8 attributes each. The province and region datasets instead contain, respectively, 107 tuples with 14 attributes and 20 tuples with 13 attributes. The attributes in each dataset are the following:

region	province	city	downloadSpeed_mbps
uploadSpeed_mbps	latency_ms	tests	stateOfWorks

Table 1: Attributes in the city-level dataset.

region	province	downloadSpeed_mbps	uploadSpeed_mbps	latency_ms
tests	Unknown	In definitive planning	In executive planning	Scheduled
Being implemented	In progress	Being tested	Done	

Table 2: Attributes in the province-level dataset.

region	downloadSpeed_mbps	uploadSpeed_mbps	latency_ms	tests
Unknown	In definitive planning	In executive planning	Scheduled	
Being implemented	In progress	Being tested	Done	

Table 3: Attributes in the region-level dataset.

3 Visualizations and Interactions

The system is presented as a web page, using the *D3.js* javascript library and CSS. There are three main visualizations: In addition, there is a fourth "visualization" describing the details of the selected data. All the visualizations are interactive and coordinated with each other: selecting something in one visualization will make it so the corresponding element is also selected in the other visualizations and added to the detail section. Furthermore, when hovering over an element, a tooltip is shown in the current visualization with details about the element and the element is also highlighted in red in all the visualizations.
Notably, each of the visualization is usable on three levels: City, Province and Region. Performing an hover or selection will act differently based on the level of the visualizations. For example, hovering over a city will only highlight the corresponding city in the other visualizations, but it will have no effect on visualizations showing an "higher" level (i.e showing provinces or region). In contrast, operating on a higher level will have effect on lower levels visualization: for example selecting a region will also select (in the other visualizations) all the provinces and all the cities inside the region.

- Choropleth map;
- Parallel coordinates;
- Scatter plot.

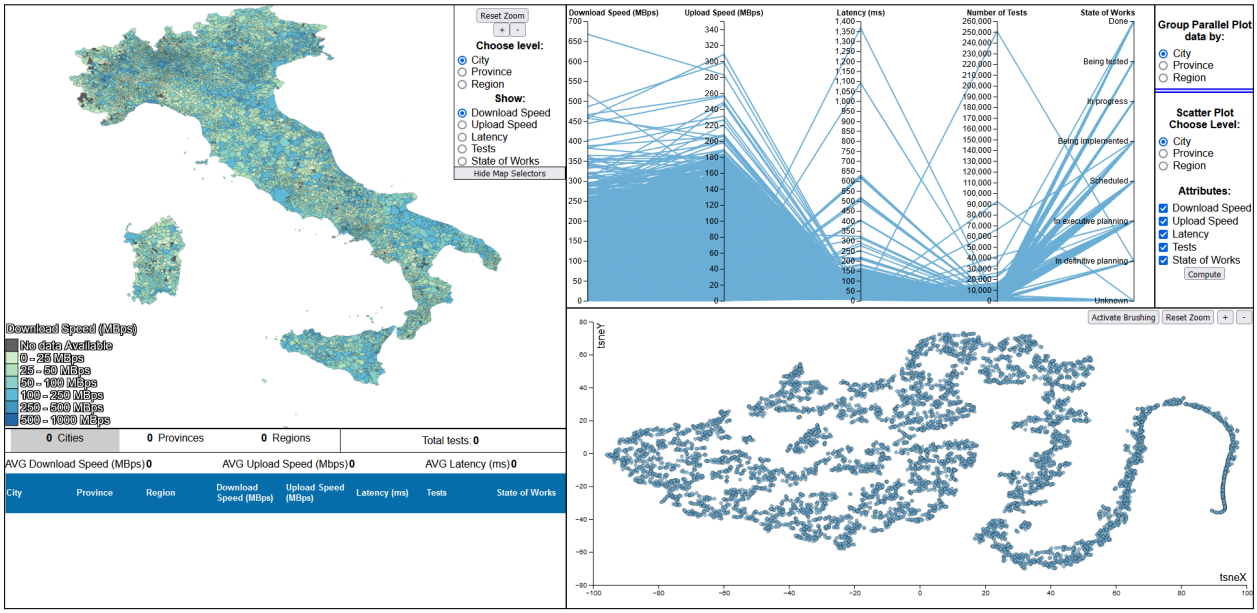


Figure 1: Screen of the visualization system.

3.1 Choropleth map

A *choropleth map* is a type of statistical thematic map where regions, states or geographical areas are colored, using different colors and/or intensities. In particular, this is a *classified* choropleth map, where the range of values is separated into classes, with all of the districts in each class being assigned the same color. The map is drawn using a GeoJson file. In particular, since the choropleth map can be viewed on three different levels, one file for each level was used.
The user can select the level and what attribute should be used to color the map. The attribute used to color the map can be chosen amongst the numerical ones (Download speed, Upload speed, Latency, Number of tests) or, if the "city" level is selected, the user can also choose to color the

map with the categorical attribute "state of works". The colors used for the attributes were taken using the Colorbrewer2 website: in particular, for the numerical attributes the colors used are a multi-hue sequential color scheme, whereas for the categorical attribute a qualitative color scheme is used.
The last element in the visualization is the legend, positioned in the bottom left corner, which describes the meaning assigned to each color. Notably, the UI related to the choropleth map can be hidden/shown with a button (except for the legend which is always visible).
It must be noted that there are some cities in the map not associated to any data in the dataset. This is evident to the user as they are colored in grey, and it is also reflected in the legend with a "No data Available" label.

3.1.1 Interaction

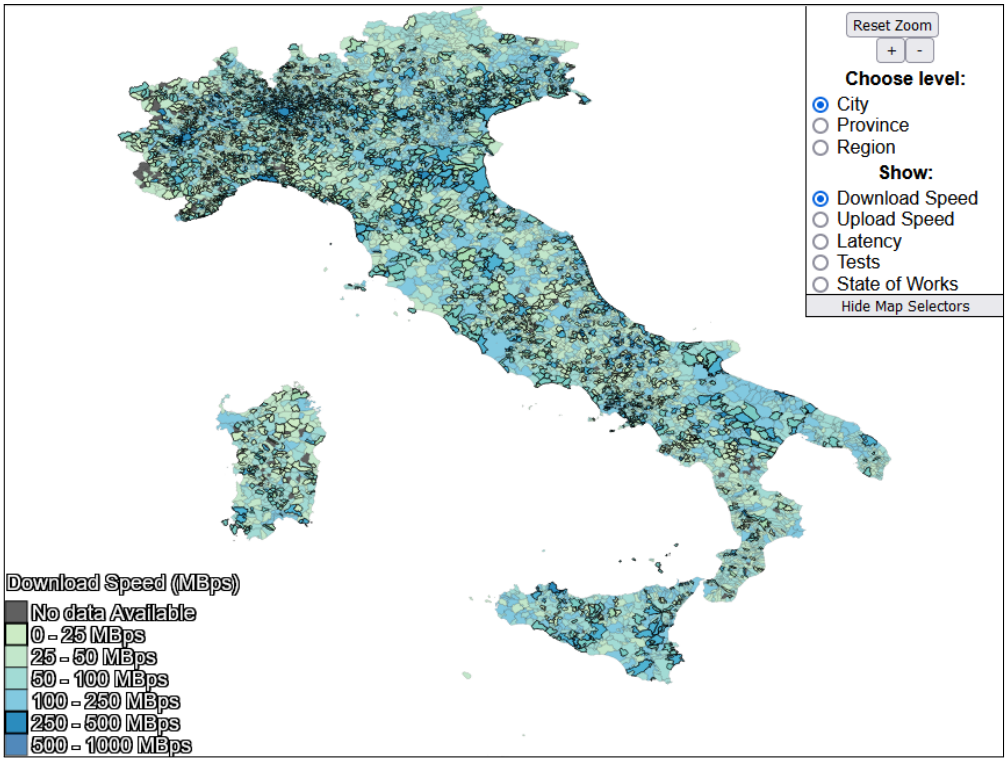


Figure 2: Example of the Choropleth map after having selected some elements.

The choropleth map is associated to a hideable UI which can be used to select the level and attributes used to color the map. In addition, the user can zoom/pan over the map using his mouse. He can also zoom or reset the zoom using buttons contained in the hideable UI.
Whenever the user hovers over an element in the map a tooltip appears, showing all the details of the hovered element. For example, hovering over a city will show the city's name, province, region, download speed, upload speed, latency, number of tests and current state of works. As previously said, hovering over an element will highlight the corresponding element in the other visualizations. The user can also select elements on the map: clicking on an unselected element will select it, whereas clicking on a selected element will deselect it. In addition, (unless the current level is "city"), all the lower level elements will be selected: for example, selecting a province will select all the cities inside that province. There is also the possibility for the user to select elements by interacting with the legend: clicking on one of the legend squares will select all the elements related to the chosen range.
Selected elements are highlighted on the map: the selected elements' opacity increases, while the others have their opacity decreased. In addition, the legend squares are similarly highlighted if all

the elements related to each square are currently selected.

3.2 Parallel Coordinates

Parallel coordinates are used to visualize high-dimensional datasets, allowing to compare and analyze many variables together. In this visualization, each attribute of the dataset is assigned to an axis and all of the axis are placed nearby each other in parallel. Each entry in the dataset is visualized as a broken line, connecting together one point for each axis representing the value of the entry for each attribute. The parallel plot has 5 axis when showing the city level: Download speed, Upload speed, Latency, Number of tests and State of Works. When showing a different level, the State of Works axis is removed, leaving only the first four.

3.2.1 Interaction

The user can select the level of the visualization with the UI on the right. To save space, the Parallel Plot UI is shared with the Scatter Plot UI. However, they are separated by a colored line.
Brushing is the main interaction the user can perform on the parallel coordinates: by clicking and

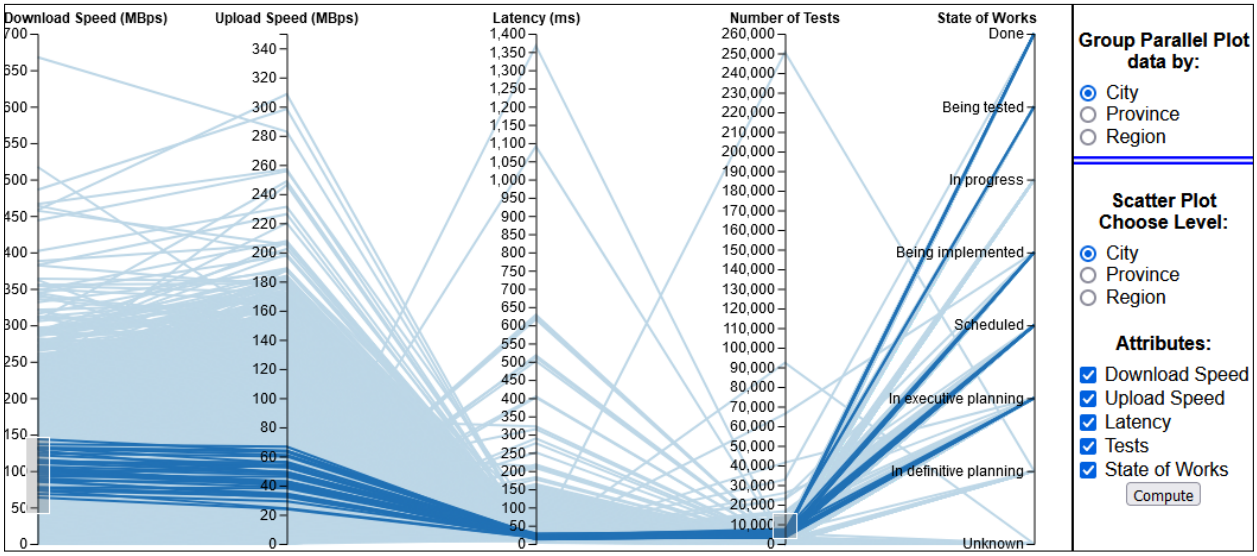


Figure 3: Example of the Parallel coordinates map after having applied some filters.

then dragging on one of the axis, the user can apply a filter on the range of values for a specific attribute. One filter can be applied for each axis, but multiple filters can be applied together. If no brushing is applied over an axis, then it is considered as if the user has applied no filter for that specific attribute. If multiple filters are applied, the only elements filtered are those satisfying all the filters. The user can remove a brush by clicking on the axis where there is no brush.
Notably, the main issue with parallel coordinates is overplotting: this happens when the dataset represented is too large, resulting in a very high number of lines overlapping with each other. In order to reduce this effect, whenever a filter is applied the filtered lines are highlighted: they become darker and their opacity increases, whereas lines out of the filters become lighter and more transparent.
The user can also manually select single elements in the dataset by clicking on the specific line. Finally, the parallel coordinates is more thought as a "filter": for this reason, whenever the user creates a new brush, all the current active selections are removed and only the elements satisfying the new filters are applied. Lastly, recreating the parallel plot by changing level resets all the current selections.

3.3 Scatter Plot

The *Scatter plot* is a type of plot using Cartesian coordinates to display values for two variables for a set of data. In this project, the scatter plot allows to visualize the results of the dimensionality reduction applied to the dataset. The dimensionality reduction used is *t-Distributed Stochastic Neighbor Embedding (t-SNE)*. The UI for the scatter plot is not positioned next to the visualization but instead it is above it, together with the parallel plot UI, as shown in Fig. 3.
In the scatterplot, every element is plotted as a circle. In particular, the user can choose on the UI what elements should be considered to perform the t-SNE, which will then be computed in real time. The x and y coordinates of each circle will then be computed based on the results. In particular, when selecting more than two attributes, the t-SNE will actually be computed and therefore the X and Y values will match the values of the t-SNE first two columns. Instead, when selecting only two attributes, the X and Y values will be computed based on the element attribute values. However, if one of the two attributes is "State of Works" for a level higher than city, the t-SNE will be computed. Indeed, each province (or region) does not have a single state of works, but it only has a count of how many cities inside it are currently in each state of works. Therefore, after normalizing the number of states, t-SNE is computed.

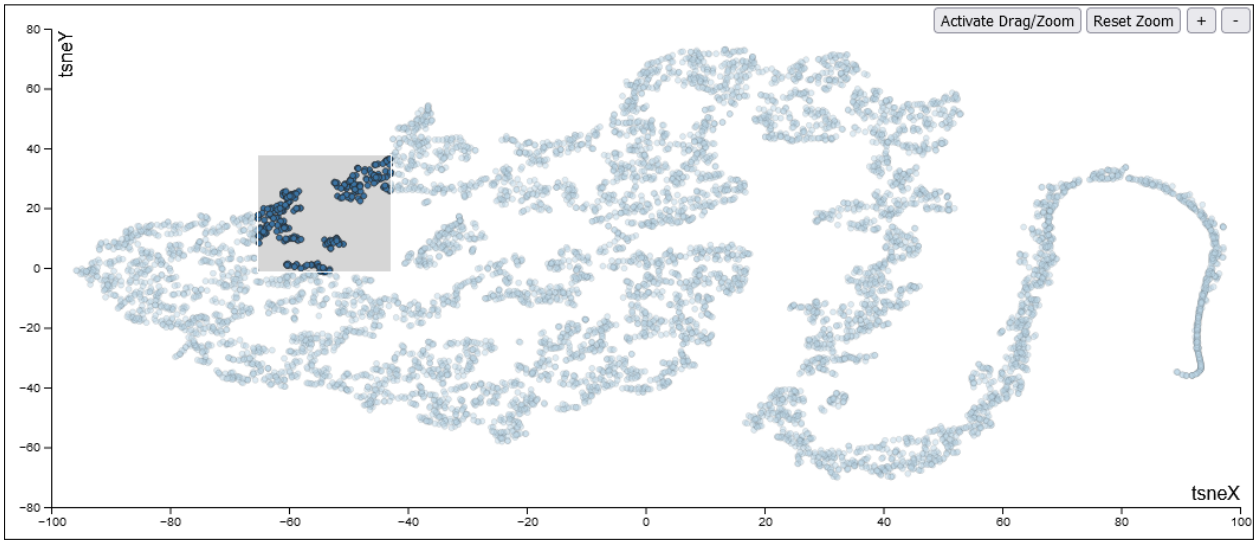


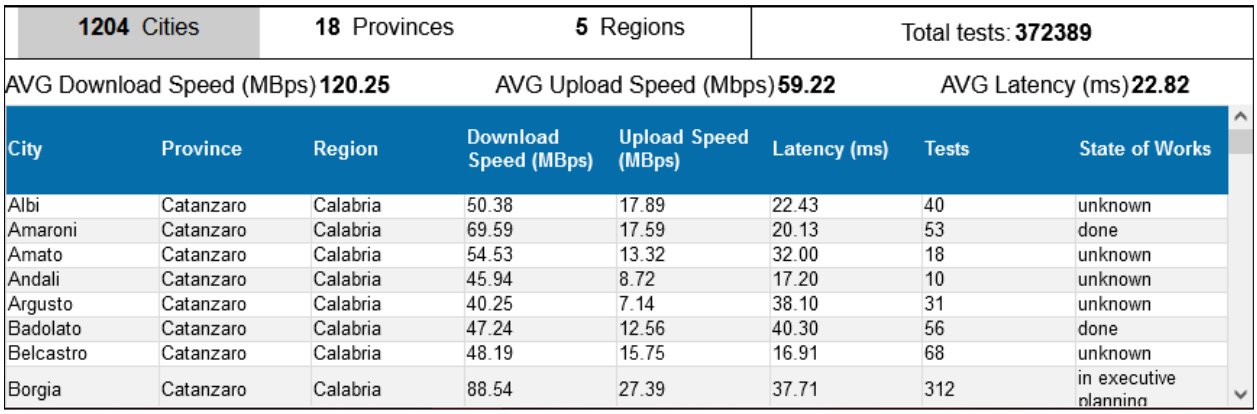
Figure 4: Example of the Scatter plot after having applied Brushing.

3.3.1 Interaction

The user can perform zoom and drag operations over the scatterplot. There are also buttons to change the zoom level and to reset it. Another button allows to switch between the Brushing and drag/zoom functionalities. Brushing lets the user apply a filter on the elements by clicking and dragging over the scatterplot. To remove the brush the user needs to click outside of it. In addition, (when the brushing functionality is disabled) the user can manually click on a circle to select the element. Doing so does not remove the brushing, so it is possible to apply a filter first and then add to the selection any element of interest. Whenever a brushing is performed/deleted or the scatterplot is recomputed, the current selection is reset.
Just like in the other visualizations, hovering over an element will highlight it in red in the other visualizations and generate a tooltip describing the element.

3.4 Detail Section

The last "visualization" in the system is a Detail section. This section contains a recap of all the elements selected by the user: it shows the total number of cities, provinces and regions selected by



City	Province	Region	Download Speed (MBps)	Upload Speed (MBps)	Latency (ms)	Tests	State of Works
Albi	Catanzaro	Calabria	50.38	17.89	22.43	40	unknown
Amaroni	Catanzaro	Calabria	69.59	17.59	20.13	53	done
Amato	Catanzaro	Calabria	54.53	13.32	32.00	18	unknown
Andali	Catanzaro	Calabria	45.94	8.72	17.20	10	unknown
Argusto	Catanzaro	Calabria	40.25	7.14	38.10	31	unknown
Badolato	Catanzaro	Calabria	47.24	12.56	40.30	56	done
Belcastro	Catanzaro	Calabria	48.19	15.75	16.91	68	unknown
Borgia	Catanzaro	Calabria	88.54	27.39	37.71	312	in executive planning

Figure 5: Example of the Detail section.

the user. In addition, it shows the average speeds and latency values and the total number of tests of the elements selected. The last element inside this section is a simple table which lets the user see the details of the selected elements.

3.4.1 Interaction

The user can change the level of the section by clicking on the total number of cities/provinces/region. By clicking on the name of an attribute in the table head, the table will be sorted (ascending order) by that attribute. Clicking again on the same attribute allows to sort in a descending order. Lastly, hovering over an element in the table will highlight it in the visualizations. Selections can't be performed from this section, as it is simply a recap of what has already been selected.

4 Analytics

4.1 Map the Ookla speedtest data to the corresponding geographical element

The ookla speedtest data is aggregated in tiles, each of them associated to a geometry but without any geographical meaning. As a consequence, by only looking at their dataset it is impossible to know what city/province/region is the tile in. The python library *Geopandas* was used to solve this issue: given a GeoJson of Italy (with all the cities), Geopandas allows to assign each tile to the corresponding Italian city. This operation also removes all the tiles not in Italy.
After assigning each tile to an Italian city, the tiles geometries are no longer needed (as the drawing of Italy in the Choropleth map directly uses the GeoJson file), therefore the subsequent operations can be performed with the python library Pandas: given the tiles, a groupby operation can be performed, grouping the tiles per city, province and then region, creating the three datasets used in the project. When grouping the tiles, it is needed to compute the various averages of the data (weighing the average on the number of tests per tile).
In total, this preprocessing of the dataset took around one and a half hour. Because of the high computation time required, the three datasets were precomputed, in order to reduce the computation at run-time. Analogously, the system does not show the "tile" level, as its dataset is too big.

4.2 Dimensionality Reduction

The scatter plot visualization shows the result of the computed dimensionality reduction. There are many techniques for the reduction, but for this system *t-distributed stochastic neighbor embedding (t-SNE)* was used. The main reason why it was chosen is the fact that t-SNE is faster than the other techniques. Since the reduction is computed in real-time based on the inputs of the user, who can choose the attributes on which the reduction should be performed, having a reasonable computation time was essential. Using t-SNE, the computation is almost immediate when executed on the province or region level, given the fact that their datasets are quite small, and it takes around one minute when executed on the city level.

4.2.1 t-SNE on the State of Works attribute

At city level, the state of works attribute has a string value assigned, namely *"Unknown", "In definitive planning", "In executive planning", "Scheduled", "Being implemented", "In progress", "Being tested"* or *"Done"*. This makes it so that t-SNE can't be directly performed using this attribute because it does not know what each string means. To deal with this issue, before executing the t-SNE the string value is converted to a number, since the string can be roughly considered as "percentage of progress in the state of works".
After the conversion, the "Done" state is assigned the highest number (7), whereas "Unknown" is assigned the lowest (0), since an "unknown" state means that there are no plans to do anything in the city. After replacing the numbers, the t-SNE can be performed.

Unknown	In definitive planning	In executive planning	Scheduled	Being implemented	In progress	Being tested	Done
0	1	2	3	4	5	6	7

Table 4: Mapping the "state of works" attribute to numerical values.

At province or region level there isn't directly a "state of works" attribute. Instead, each element is assigned more attributes containing the total number of cities currently being in each state of works. In this case, the value is not a string and therefore it would be possible to directly perform the t-SNE. However, the values need to be normalized first, therefore the t-SNE is actually performed when these attributes contain the percentage of cities currently being in each state of works.

5 Related Works

5.1 Ookla

Ookla, being the source of the dataset used in this system, also produced a similar visual analytics tool based on an interactive heat map, [6]. The map is the only element in their visualization tool: it is directly based on their dataset, directly showing the tiles drawn over the world. Their system offers a higher level of detail, allowing to zoom in and see exactly the specific tiles. However, the tiles are not geographically assigned, therefore it is not possible to exactly know what city each tile belongs to. Finally, their system has a more generic scope, allowing to see data related to the whole world.

5.2 BUL

The BUL plan offers a web page [3] with a choropleth map showing the current state of works for each city, similarly to what has been done in this system. The map also adds some more details, such as the specific fiber track, the schools connected with fiber and the single housing units. In addition, it also offers a more generic "region level" view, with the choropleth map's coloured shades showing the percentage of completed Fiber working sites. Contrarily to this system, it does not have a "province level". Contrarily to the Ookla's visualization system, the tool is not only limited to the map but it includes more features describing details of the works. On a final note, the BUL map does not contain any info related to the internet connection values (download speed, upload speed or latency), being only focused on the informations related to their project.

5.3 Other works

While the platform for the BUL data does serve a similar purpose (related to the BUL informations) to the system presented in this report, I was unable to find any other project whose objective was to produce a visualization system with a similar objective of showing internet data. There are many papers in which the authors describe the impact of the diffusion of broadband internet, on the economy, society, culture and more (for example [7]). However, these papers are focused on the impact of the diffusion of high-speed, high-quality internet, leaving little room for the discussion of the diffusion itself.
Regarding the analysis on the impact of the diffusion of high speed internet, the platform described in this report can prove useful: if one has some data related to various cities in Italy (for example the percentage of people who work from remote, or the number of industries per city), this system might aid in the analysis of a correlation between the data of interest and internet speeds. Possibly, the system's dataset could be combined with the dataset of interest, slightly modifying this system in order to be able to analyze everything on this platform, similarly to what was done with the BUL data.

6 Insights

This system has been designed with multiple intended users in mind:

- The system user can be someone looking to gather insights on the Italian internet characteristics. This could be done for multiple reasons, ranging from writing a journalistic article to research purpose.
  Suppose for example that a reporter wants to write an article on the top and worst speeds in Italy. Suppose aswell that the reporter wants to ignore the cities with a very low number of tests in order to remove the smallest cities and any potential outlier: he can use the parallel coordinates "Tests" axis to filter only the cities with at least 500 tests. Then, thanks to the Detail section, he can sort by "Download Speed". He's then able to find the top 10 and worst 10 cities in Italy:

City	Download Speed (MBps)
Grugliasco	311.08
Beinasco	305.62
Settimo Torinese	295.20
Sesto San Giovanni	292.96
Venaria Reale	292.06
Torino	290.60
Milano	287.39
Collegno	287.30
Torrevecchia Pia	287.28
Nichelino	285.09

City	Download Speed (MBps)
Fanano	20.77
Gualdo Cattaneo	22.20
Bagni di Lucca	25.86
Ville di Fiemme	28.27
Sant'Agata de' Goti	29.01
Ferno	30.10
Prignano sulla Secchia	30.52
Amelia	30.81
Veronella	30.94
Livigno	31.39

Table 5: Top 10 and Worst 10 Cities for Download speeds (MBps) (for cities with at least 500 tests).

- As previously mentioned in 5.3, a possible system user could be someone trying to understand whether there exists a correlation between some italian data and the quality of the internet in the cities.
  For example, suppose that the user has a dataset indicating some similarities betweeen a few cities. He can use the dimensionality reduction results, shown on the scatterplot, to find whether this cities are also similar in terms of internet quality. After computing the dimensionality reduction based on his attributes of interest (the reduction could be performed without using the "State Of Works" and "tests" attributes as the user might find them uninteresting), he can then use the choropleth map to manually select every city he wants to analyze, and then check on the scatter plot whether these cities are actually close or not, helping him understand whether his hypothesis of a correlation between the internet quality and his data is true or not.

- Another possible user can be a manager of some company trying to understand the current internet quality (based on the Ookla dataset) and if/how it might improve in the future (by looking at the BUL data), for example to find the best city in a certain area where to place a new division of the company.
  Suppose for example that the company needs to place a new division in the province of

Catanzaro (CZ) in Calabria. The new division does not need extremely high download/upload speed, but it is necessary that the latency is very low. Therefore, the manager uses the choropleth map (open at province level) to select the province of Catanzaro. Thanks to the parallel plot, he can immediately see that there are cities in the province with a low latency. Therefore, he can change the choropleth view to the city level, showing latencies. Afterwards, using the choropleth map he can further examinate what cities satisfy his requirements thanks to the different colours. In addition, the map allows him to see where this cities are, enabling him to verify additional requirements (for example, it might be needed that the city is nearby the sea).

- The system could prove useful to managers of a telecommunication company or of the Public Administration to understand how the BUL investments are going: in particular, being able to directly compare the BUL data with the internet quality data allows to better understand what zones of Italy might need more investments on the fiber infrastructure.
  Suppose that a telecommunication company wants to invest in the development of a new fiber infrastructure. Knowing that the company is able to quickly plan and implement the infrastructure, they decide to invest in cities flagged as "In definitive planning", in order to have an head start over the competition. However, they also want to invest in cities where there is actually the need for a good internet connection, and not in zones where the quality is already good.
  The manager in charge of the task can use the parallel plot (or the choropleth map showing the "state of Works") to filter only the cities where the BUL state is "In definitive Planning". The other axes of the parallel plot allow to filter the cities with low internet quality: for example, download speed lower than 70 MBps, upload speed lower than 20 MBps. In the detail section, he can then find the best 5 cities where to invest:

City	**Province**	**Region**	**Download Speed (MBps)**	**Upload Speed (MBps)**	**State of works**
Acri	Cosenza	Calabria	49.54	14.09	in definitive planning
Campagna	Salerno	Campania	52.91	14.90	in definitive planning
Lierna	Lecco	Lombardia	41.31	10.67	in definitive planning
Novate Mezzola	Sondrio	Lombardia	63.73	13.93	in definitive planning
Santa Maria a Vico	Caserta	Campania	57.72	15.95	in definitive planning

Table 6: Research results (sorted by city name) of the example described.

7 Conclusions and Future Works

The visualization system presented offers an interactive tool to analyze and gather insights on the internet quality in Italy and the development of works on the infrastructure. The user has various tools at his disposal to perform the analysis, hastening the process. The tools allow him to compare the various zones of italy amongst each other, finding similarities and dissimilarities. The filtering capabilities enable the user to quickly find elements in the dataset satisfying their requests (or to find out whether such elements do not exists).
Despite having interactive and coordinated visualizations which might make the system appear complex at a first glance, in the end it is user-friendly, resulting usable even by a non-so-experienced user.
The Ookla datasets are updated every quarter of the year, and this system uses the latest published (Q1 2023). Overtime, the dataset should be updated, together with the official Ookla's. Eventually, it could be interesting to look, using the platform, at how the data changes overtime.
Given the lack of similar projects, this system could be further modified, catering it to the specific needs of a single user, eventually combining the dataset with others, in a similar fashion to what was done with the BUL data. One particular improvement that could be made would be to combine the Ookla dataset with official datasets from the telecommunication company, allowing to make a direct comparison.

8 References

References

[1] Banda ultra larga. Accessible at `https://bandaultralarga.italia.it/en/`.

[2] Bul dataset. Accessible at `https://bandaultralarga.italia.it/wp-content/uploads/stato_lavori.csv`.

[3] Bul map. Accessible at `https://bandaultralarga.italia.it/en/map/`.

[4] Ookla's open data github page describing their dataset. Accessible at `https://github.com/teamookla/ookla-open-data`.

[5] Ookla's open data initiative. Accessible at `https://www.ookla.com/ookla-for-good/open-data`.

[6] Ookla's open data initiative interactive map. Accessible at `https://www.ookla.com/ookla-for-good/open-data#interactive-map`.

[7] P. Koutroumpis. The economic impact of broadband on growth: A simultaneous approach. *Telecommunications policy*, 33(9):471–485, 2009.