Hafshah Ashrawi
Yidan Chen

ADDRESSING DIGITAL GAP IN NYC USING NYCLINK AND CENSUS DATA

A. Introduction

According to the 2015 report of NYC Bureau Policy, approximately a quarter of NYC households and 32% of unemployed people lack the broadband internet at their home (NYC Bureau Policy Report, 2014). As a practical solution to this issue, the Mayor of New York in 2014 launched an initiative by partnering with CityBridge called LinkNYC. LinkNYC uses kiosks ("Links") to provide internet connectivity to the five boroughs (Manhattan, Brooklyn, Queens, the Bronx, and Staten Island) and is said to give a new use for the City's unused payphones. Five years later, CityBridge had launched around 1.774 of 7500 planned wifi-kiosks or more popularly known as LinkNYC. Citybridge itself is a consortium which consist of CIVIQ Smartspaces, Qualcomm, and Antenna, they are responsible for designing and maintaining the network of Link NYC.

The main idea behind this project was to create an open, equal and connected city (Huber, 2016 in vice.com). Several sources also emphasize the goal of this project to bridge the 'digital gap' and cater to the need of those people really in need of free public wifi services. However, some findings denote that the wifi kiosks are mostly located in Manhattan where there are fewer people, particularly in demand for this kind of services (Correal, 2019 in Newyorktimes.com). Also, less than 10% of planned LinkNYC kiosk is to be installed in Bronx where there is the highest percentage of households without internet access (NYC Bureau Policy Report, 2015 and Huber, 2016 in vice.com). Hence, it is interesting to analyze just how central the 'digital gap' is to the mission and the real implementation of this project. Finally, in this essay, we would like to identify how this project has been catering to particular people in need of this service.

After defining our aim of this project, we break down our goal to several sub-questions which can help us sequence analyzing the process. The main questions for this project comprise of :

1. What is the principal borough where most public wifi services are located?
2. Between the five boroughs in NYC, where is the pocket of poverties in NYC mostly concentrated?
3. Is there any statistical or spatial correlation between poverty/income-level and public wifi location?

B. Data Source

we mainly worked with two datasets: NYC Wifi hotspot locations and 2010 census tract data. The NYC Wifi hotspot location was derived from NYC Open data portals. As for the Census data, the dataset was taken from American community survey and pertains to 2010 census. Both datasets were downloaded as CSV files and were processed using Numpy, Pandas, and ArcGIS. The NYC Wifi Hotspot location contains information of all the publically accessible Wifi. Information includes the location of the Wifi kiosks, providers, types (free/limited free), and activated time, while the census data include information regarding population, poverty level, race, income level, and census tract is used as the geographical unit, and geometry point.

C. Methodology and Results

We distinguished the step of this project in to data cleaning, data analysis, and data visualization. The detailed descriptions of each step will be explained in the next section

Data Cleaning

1. Our first step is to eliminate the unnecessary column in NYC Wifi dataset such as latitude, longitude, geometry point, type of kiosk, etc. The NYC Wifi data also has a column pertain to borough code and census tract code. This code is a sub-part of census tract code which also available in the Census Tract dataset. Another thing to consider is that the NYC Wifi dataset contains multiple census-tract codes. For example, in "BOROCT2010" column, there are four rows of ' 3118400' values which translate that there are four public wifi-kiosk in the '3118400' tract. We address these issue by making two separate datasets. **First, we make a new dataset which contains a new column.** The new column will represent the **new census tract**

Picture 2. Data Cleaning Part for Census Data

```
import csv

datastream = csv.reader(open('output2','r',encoding = 'utf8',errors = 'ignore'), delimiter = ',')
CSVheaders = datastream.__next__()

boroughs = {'Manhattan':'36061','Bronx':'36005','Brooklyn':'36047','Queens':'36081','Staten Island':'36085'}
clean_publicwifi = {}
ctracker = []

#store individual census tract by new identifier based on other census file, boroughs + ct_2010
for line in datastream:
    sline = {CSVheaders[i]:line[i] for i in range(len(CSVheaders))}
    newID = boroughs[sline['BORONAME']] + sline['BOROCT2010'][1:]

    #If there is multiple FIDs within a census tract
    if newID not in ctracker:
        if sline['PROVIDER'] != '-' and sline['PROVIDER'] != '0':
            #only keeping relevant datanew
            clean_publicwifi[newID] = {'BOROCT2010':sline['BOROCT2010'],'location':[sline['LOCATION']],'Provider'
            ctracker.append(newID)
    #if census exist,append the provider list
    else:
        if sline['PROVIDER'] == '-' and sline['PROVIDER'] == '0':
            clean_publicwifi[newID]['Provider'].append(sline['PROVIDER'])
            clean_publicwifi[newID]['location'].append(sline['LOCATION'])
```

```
dataframe_publicwifi = pd.DataFrame.from_dict(clean_publicwifi).transpose()
dataframe_publicwifi1 = dataframe_publicwifi.reset_index()
dataframe_publicwifi1
```

code that similar with the id in census tract dataset. Werun some system in python to make the new column representing a set of number which composed of borough code and the last five number of each value in ‘‘BOROCT2010’.

From the US Census website, we were informed that the first 5 number of the census tract is the code for each borough. For example, Manhattan code is 36061. Therefore, the new ID in NYC Hotspot dataset, which has a value of Manhattan in ‘BORONAME’ column and 3118400 in ‘BOROCT2010’ column is 36061118400. Using this method, later we can join the NYC Wifi dataset and the census tract dataset. **Second**, we created the new dataset which contains a new column containing a total of public wifi for each census tract. We use group_by method in Pandas to make the original dataset.

2. The census data already contains an essential element for our analysis. There is no significant issue, such as a missing value. After creating a new dictionary which uses the census tract ID (similar to the unique ID in Public Wifi dataset), the dictionary then was exported to data frame to make the analysis process more straightforward. The following step is to merge the new data frame of NYC Public wifi which contains the new ID and the census tract data frame, and we rename the key-column to ‘index’ to merge both data frames. To make it more comprehensive for our analysis, we join the newly created data frames with the second data frame in Step 1, which contains the number of public wifi for each census tract. The final dataset from this data cleaning process is shown in Picture 1 below, showing that there are approximately 756 rows and 16 columns in this new dataset.

	index	BOROCT2010	location	Provider	ct2010	borough
0	36047118400	3118400	[3386 FULTON STREET]	[LinkNYC - Citybridge]	184	Brooklyn
1	36047119800	3119800	[62 PENNSYLVANIA AVENUE]	[LinkNYC - Citybridge]	198	Brooklyn
2	36047036700	3036700	[2493 FULTON STREET]	[LinkNYC - Citybridge]	367	Brooklyn
3	36047081600	3081600	[3723 CHURCH AVENUE]	[LinkNYC - Citybridge]	816	Brooklyn
4	36047085600	3085600	[3402 CHURCH AVENUE]	[LinkNYC - Citybridge]	856	Brooklyn
...	...	...	...	...	...	...
764	36061004900	1004900	[Spring St (C,E)]	[Transit Wireless]	49	Manhattan
765	36081115100	4115100	[155th St between 29th Ave and 32nd Ave]	[SPECTRUM]	1151	Queens
766	36061003001	1003001	[145 ORCHARD STREET]	[LinkNYC - Citybridge]	30	Manhattan
767	36047079000	3079000	[1339 FLATBUSH AVENUE]	[LinkNYC - Citybridge]	790	Brooklyn
768	36047056100	3056100	[817 Manhattan Ave]	[LinkNYC - Citybridge]	561	Brooklyn

769 rows × 6 columns

Picture 2. Data Cleaning Part for Census Data

```
#store information for every census tract
censusdata = {}
#store only census tracts with complete information
cleancensusdata = {}

relevantinfo = ['TotalPop','Men','Women','Hispanic','White','Black','Native','Asian','Poverty']

for line in datastream:
    sline = {CSVheaders[i]:line[i] for i in range(len(CSVheaders))}

    #keep all original data
    censusdata[sline['CensusTract']] = sline
    #use counter to make sure every variable has information, or else might skew results if not all variables
    counter = 0
    for info in relevantinfo:
        if sline[info] != '':
            counter += 1

    #this means all variables are not empty
    if counter == 9:
        cleancensusdata[sline['CensusTract']] = {}
        for info in relevantinfo:
            cleancensusdata[sline['CensusTract']][info] = sline[info]
cleancensusdata
```

```
#convert the dictionary to a dataframe.
dataframe_census = pd.DataFrame.from_dict(fixedcensusdata).transpose()
dataframe_census1 = dataframe_census.reset_index()
dataframe_census1
```

	index	TotalPop	Men	Women	Hispanic	White	Black	Native	Asian	Poverty
0	36005000200	5403.0	2659.0	2744.0	0.7508	0.0203	0.1600	0.0000	0.0402	0.2000
1	36005000400	5915.0	2896.0	3019.0	0.6207	0.0306	0.3007	0.0000	0.0003	0.1302
2	36005001600	5879.0	2558.0	3321.0	0.6501	0.0106	0.3204	0.0000	0.0000	0.2603
3	36005001900	2591.0	1206.0	1385.0	0.5504	0.0900	0.2900	0.0000	0.0201	0.3701
4	36005002000	8516.0	3301.0	5215.0	0.6101	0.0106	0.3101	0.0003	0.0303	0.5302
...	...	...	...	...	...	...	...	...	...	...
2120	36085030301	4895.0	2371.0	2524.0	0.3007	0.4002	0.1106	0.0000	0.1600	0.0807
2121	36085030302	6279.0	3093.0	3186.0	0.3508	0.2807	0.1706	0.0000	0.1403	0.1900
2122	36085031901	2550.0	953.0	1597.0	0.2701	0.0602	0.6004	0.0000	0.0603	0.3903
2123	36085031902	4611.0	2043.0	2568.0	0.2009	0.1407	0.6109	0.0000	0.0009	0.4102
2124	36085032300	1131.0	597.0	534.0	0.4505	0.2400	0.2907	0.0000	0.0000	0.2005

2125 rows × 10 columns

Picture 3. Final Public Wifi-Census Dataframe

index	TotalPop	Men	Women	Hispanic	White	Black	Native	Asian	Poverty	BOROCT2010	location	Provider	ct2010	borough	total_wifi
36005001900	2591.0	1206.0	1385.0	0.5504	0.0900	0.2900	0.0000	0.0201	0.3701	2001900.0	[141 LINCOLN AVENUE]	[LinkNYC - Citybridge]	19	Bronx	1
36005002000	8516.0	3301.0	5215.0	0.6101	0.0106	0.3101	0.0003	0.0303	0.5302	2002000.0	[660 SOUNDVIEW AVENUE]	[NYPL]	20	Bronx	1
36005002300	4774.0	2130.0	2644.0	0.6203	0.0002	0.3605	0.0100	0.0000	0.5402	2002300.0	[3 Avenue - 138 St [6]]	[Transit Wireless]	23	Bronx	1
36005002500	5355.0	2338.0	3017.0	0.7605	0.0105	0.1809	0.0000	0.0300	0.5005	2002500.0	[216 WILLIS AVENUE]	[LinkNYC - Citybridge]	25	Bronx	4
36005002702	4778.0	2427.0	2351.0	0.7103	0.0106	0.2602	0.0000	0.0000	0.5207	2002702.0	[710 EAST 138 STREET]	[LinkNYC - Citybridge]	2702	Bronx	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
36085019800	6690.0	3538.0	3152.0	0.0509	0.8802	0.0105	0.0005	0.0307	0.0503	5019800.0	[South parking lot Entrance]	[SPECTRUM]	198	Staten Island	5
36085020700	6178.0	3046.0	3132.0	0.5302	0.0806	0.3507	0.0000	0.0105	0.3304	5020700.0	[75 BENNETT STREET]	[NYPL]	207	Staten Island	1
36085020804	5602.0	2648.0	2954.0	0.0305	0.8508	0.0000	0.0000	0.0806	0.0708	5020804.0	[830 HUGUENOT AVENUE]	[NYPL]	20804	Staten Island	1
36085024401	6408.0	2755.0	3653.0	0.0803	0.8705	0.0000	0.0000	0.0201	0.0404	5024401.0	[7430 AMBOY ROAD]	[NYPL]	24401	Staten Island	1

756 rows × 16 columns

Data Analysis and Data Visualization

1. What is the principal borough where most public wifi services are located?

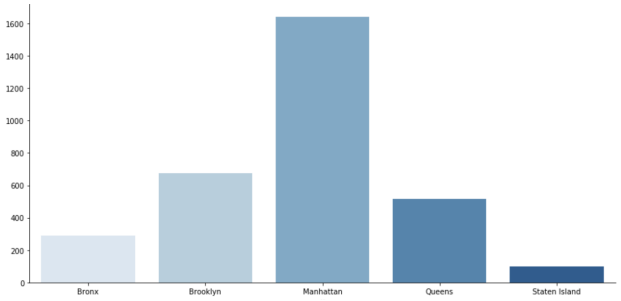
To make this graph, we use python and *seaborn* package to compute the total wifi per each borough and populate a new dictionary which contains the sum of public wifi spot for each borough. We also added spatial mapping using ArcGIS to complement the results.

Main findings:

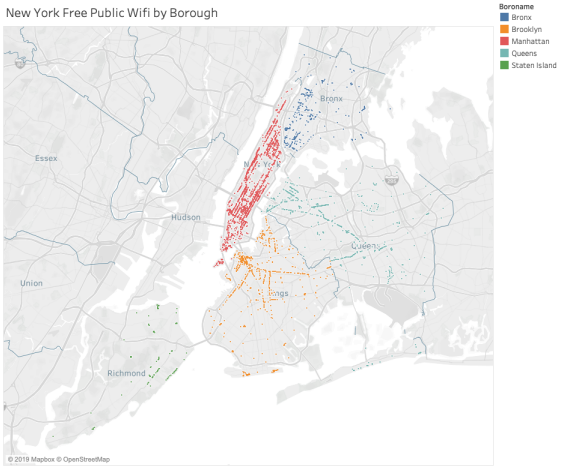
a. Manhattan has the highest number of public wifi services (1639 spot), more than double the number of public wifi in Brooklyn (673). Ranking from the highest to the lowest one, there are Manhattan, Brooklyn, Queens, Bronx, and lastly Staten Island. This findings might attribute to the ad revenue for CityBridge. It is profitable to put LinkNYC in Manhattan where most tourist visit, and a high presence of business district. According to one media source, LinkNYC expects to generate over $500 million in revenue through partnership and sponsorships, and still have to share part of the revenue to the municipal (McKetta, 2017 in Speedtest.net). Hence, we can see that for City Bridge it is really important to have a strategic location for installing wifi kiosk to boost their revenue.

b. The map also clearly shows how much of Manhattan is covered by Link, seconding the first finding. There are some areas not covered by LinkNYC for example Washington Heights North and lower Manhattan below Worth St. Links.

Picture 4. Total Public Wifi Hotspot per Borough



	Bronx	Brooklyn	Manhattan	Queens	Staten Island
publicwifitotal	291	673	1639	515	100



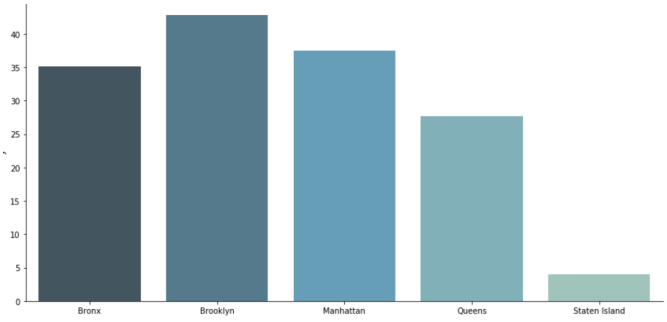
2. Between the five boroughs in NYC, where is the pocket of poverties in NYC mostly concentrated?

A similar method with the previous example was used to produce this graph. The main findings derived from the graph consist:

a. As the graph shows the percentage of low-income people out of the total population for each respective neighbourhood, Brooklyn has the highest percentage of people living under the poverty rate. Following Brooklyn (42,77%), from the second highest to the lowest one there are Manhattan (37,51%), Bronx (35,14%), Queens (27,69%), and Staten Island (3,93%) at the last.

b. The percentage of poor people in Manhattan, Queens, Brooklyn, and Bronx results in a relatively similar range (between 35% to 42%). However, it must be noted that the total population for each borough is different. For example, Manhattan' total population is 1.415.278 people, while Brooklyn is 680.164 people. In other words, Manhattan host approximately 523.652 poor people and Brooklyn' figure would be around 285.668 people. Even though Manhattan has a lower percentage than Brooklyn, it has significantly more poor people than Brooklyn.

Picture 5. Total Public Wifi Hotspot per Borough



	Bronx	Brooklyn	Manhattan	Queens	Staten Island
low income people	35.1368	42.7664	37.5135	27.6877	3.973

3. Is there any statistical or spatial correlation between poverty/income-level and public wifi location?

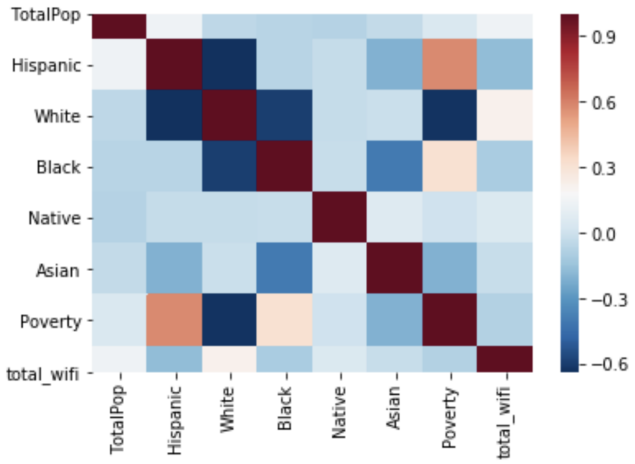
Using heatmap and scatterplot function in Seaborn, we try to map the correlation between both variables, the poverty and total public wifi. We also consider the percentage of racial class in the first correlation heatmap to broaden our findings with regards to poverty in NYC. To simplify the data frame, we also drop some irrelevant column such as provider, location, and borough code.

Main findings:

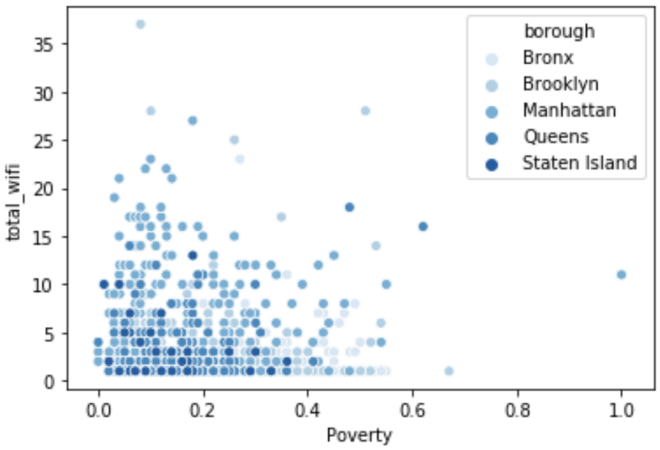
a. According to the Correlation Matrix below, the correlation between poverty and public wifi location is relatively low, suggesting the possibilities of other factors have greater influences over the location choice of public wifi. Furthermore, according to the scatterplot there is no clear pattern of neither negative or positive correlation. Looking at the dataset, the correlation figure is around -0,07 and signifies a very weak correlation. While for additional findings, we find that the correlation between poverty and hispanic class for example is quite high, while white people and poverty tend to be strongly negative.

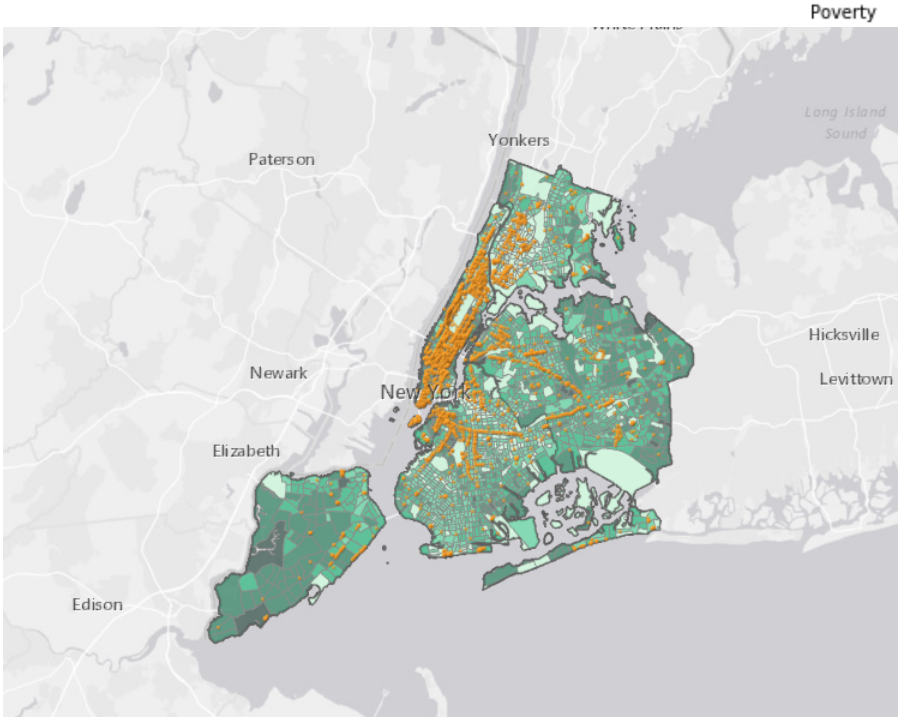
b. The map demonstrates a clear spatial correlation between the income level and the location of public wifi: higher-income neighbourhoods enjoy more access to public wifi while free public wifi is commonly absent in low-income neighbourhoods. The trend is clearest in Manhattan borough and least obvious in Staten island as only few wifi hotspots exist and income gap between census tract is not as substantial.

Picture 6. Heatmap of Correlation in NYC Wifi-Poverty Dataframe



Picture 7. Heatmap of Correlation in NYC Wifi-Poverty Dataframe





Link to access the map https://arcg.is/199muG0

D. Conclusions

The goal of this paper is to examine the digital gap in New York City- whether there is an equity issue regarding public wifi access. It is rational to place LinkNYC (public wifi kiosks) in higher-density neighbourhoods, and as a matter of fact, Manhattan does outnumber all other boroughs. However, our findings raise the issue of equity and question whether the program has managed to serve its target population- people who lack access to the internet. We find a clear spatial pattern that Link NYC are concentrated in higher-income/lower-poverty neighbourhoods and vice versa. We urge the City to address this digital gap and do more to provide public wifi access for low-income communities. Campaigns and other education programs can also be initiated to encourage more equitable uptake of free public wifi.

E. Limitations

1. Is poverty a good indicator to represent people who lack access to broadband internet at home? As a general rule of thumb, it might show some correlation but further research can complement this assumption. For example, in this project we find less correlation between poverty rate and total wifi installed, perhaps the more specific definition of poor would
2. The compliance of time attribute between two datasets. For a metropolis city such as New York, it is highly possible that the demographic composition can significantly change in a relatively short year. However, we also bumped into the fact that the comprehensive demographic data like census is only conducted once every 5/10 years.
3. Other factors may also have a high correlation with access to Wifi. For example, population density and tourists volume. Being an open, equal, connected city can include various objectives, and due to limited resources, some areas may be prioritized due to other concerns.

F. Endnote/References

Data Source

1. NYC Open Data Portal, NYC Wifi Hotspot Location. link to access: https://data.cityofnewyork.us/Social-Services/NYC-Wi-Fi-Hotspot-Locations/a9we-mtpn?fbclid=IwAR1YLzOXETv-pxHqtK5d8zwe1JZGI_mow4h67ot9s0SBfe4abRiULXQlpwI
2. US Census Data. link to access: https://factfinder.census.gov

Reference

3. NYC Bureau Policy Report. (2014).”Policy brief: Internet Inequality, Broadband Access in NYC”. Office of the New York City Comptroller. Bureau of Policy and Research. link to access: https://comptroller.nyc.gov/wp-content/uploads/documents/Internet_Inequality.pdf
4. Huber, L. (2016, September 14). “Is New York City’s Public Wi-Fi Actually Connecting the Poor?” in www.vice.com. link to access: *https://www.vice.com/en_us/article/vv7pw3/linknyc-is-bringing-internet-to-new-yorks-most-disconnected-people*
5. Correal, A. (2019, December 6). “Just a Quarter of New York’s Wi-Fi Kiosks Are Up. Guess Where.” in www.newyorktimes.com. link to access: https://www.vice.com/en_us/article/vv7pw3/linknyc-is-bringing-internet-to-new-yorks-most-disconnected-people
6. McKetta, I. (2017, October 31).”LinkNyc Proves Public Wi-Fi Can Be Free, Fast And Far-Reaching” in www.speedtest.net. link to access : https://www.speedtest.net/insights/blog/linknyc-2017/