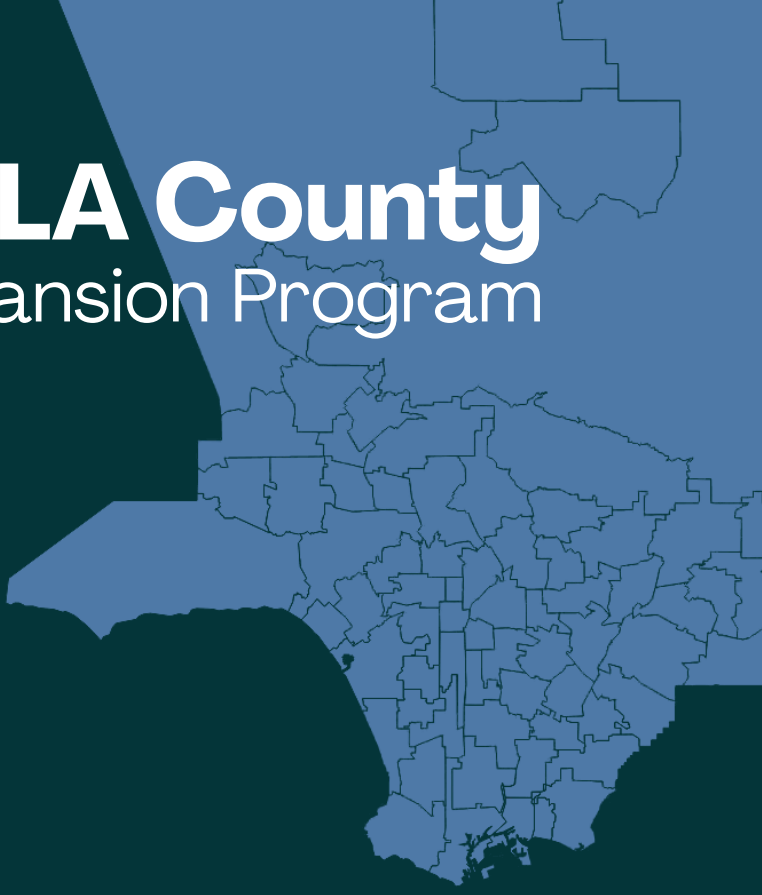
LA County
Internet Expansion Program



TEAM #104 | SPRING 2021


Francisco Razo Gomez


Gerardo Mora


Marlene Plasencia


Esparta Palma


Jassmine Rabii

Table of Contents

Introduction

How can Los Angeles County use its existing infrastructure to bring Wi-Fi to those most in need of broadband access?

PROBLEM

According to the Federal Communications Commission, nearly 30 million people in the US do not have access to high-speed fixed internet service. With the rise of work-from-home and more than ~5 million school-age children pushed into online learning, this inability to access online resources has led to achievement gaps in the community which potentially lead to higher income and education disparities down the line.

COVID-19 restrictions have closed down commonly used indoor spaces like cafés and libraries where people normally obtained broadband access. To mitigate the growing inequality gap the pandemic has caused, local governments need to increase internet access by leveraging existing community infrastructure.

HYPOTHESIS

Identifying public resources which can improve coverage will help Los Angeles County advocate for more reliable public zones of internet access that can be established to reduce the internet resource gap that communities experience.

Through recognition of internet dry spots, we can empower Los Angeles county and other similar communities to expand their internet access to a larger geographic area and close the digital divide.

Data

Datasets will enable us to analyze the map locations of bus stops, education, and income disparities, and overall need-based internet access within LA Counties.

Datasets

LA County: Bus Stops

Source: https://geohub.lacity.org/datasets/bus-stop-benches/
Size: ~1.5 MB
Description:
Provides up-to-date bench location data, type, the city which we can use to cluster our results on a map. Each bus stop will represent a point on the map which will serve as a location to view if our target demographic is representative within its area. The API can be downloaded in its entirety with possible format options. The downside is that the data does not provide street names, but can be inferred with a reverse lookup using lat/lng.

LA County: Public Wifi

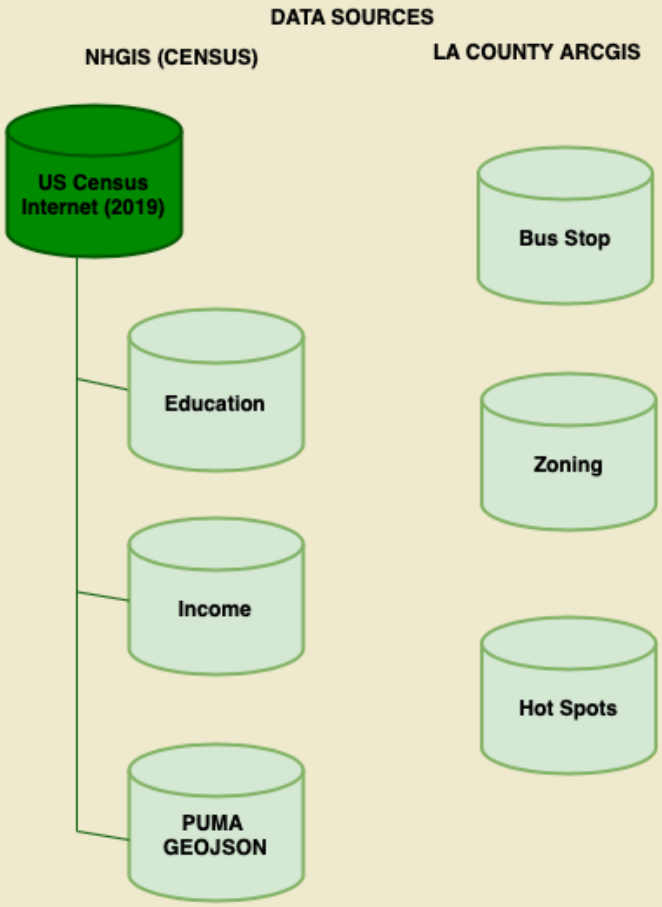
Source: https://lacounty.maps.arcgis.com/apps/mapviewer/index.html?layers=c9355c8acef5427d9e110990f2e98ebb
Size: ~4 KB
Description:
LA County has a map of all public and commercial data points of existing WiFi. This will enable us to not duplicate the WiFi service in a specific area where it already exists. The upside to this data set is we can use it as a resource for users to search for existing services on the map. The issue with the data is that we do not reliably know if these devices are actually in working order or if they are still accessible.

LA County: Zoning

Source: https://geohub.lacity.org/datasets/zoning
Size: ~1.2 MB
Description:
Zoning data can be used later on to prioritize potential infrastructure at certain bus stops like commercial vs residential. Part of our takeaway conclusions.

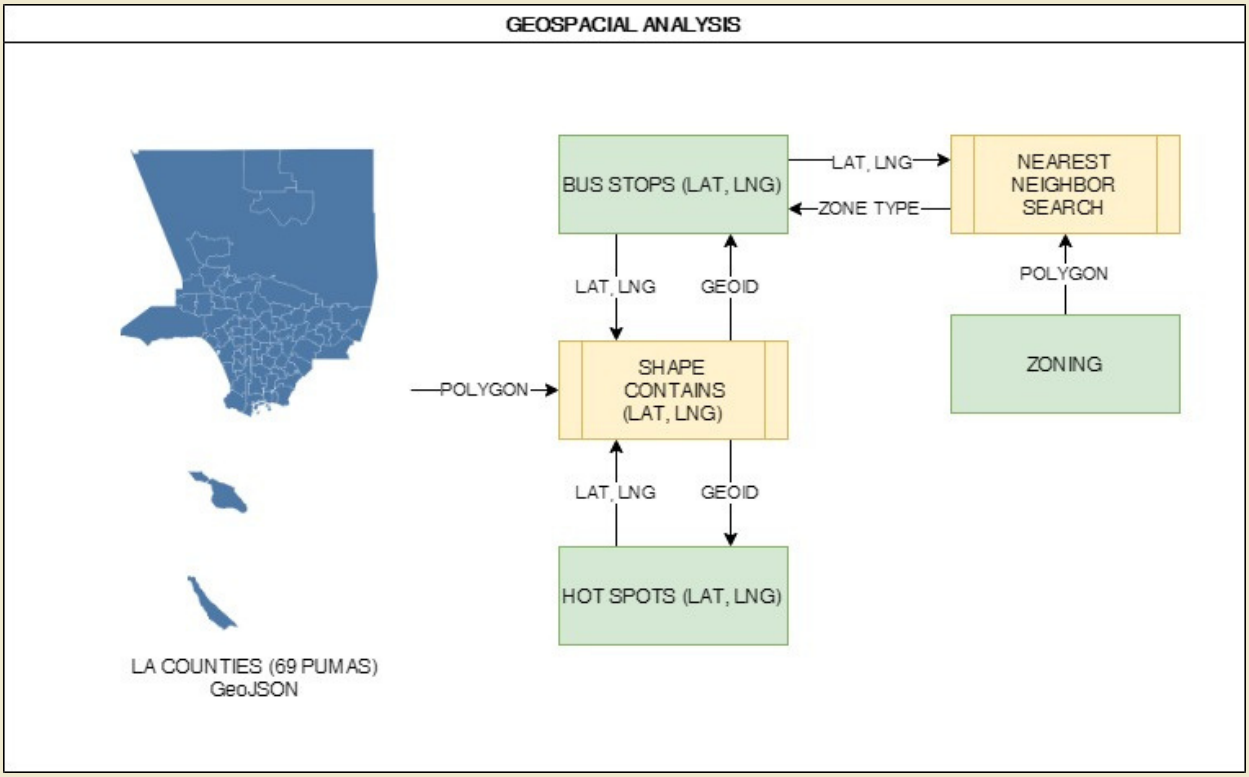
NHGIS | Census | 2019_ACS1

Source: https://data2.nhgis.org/main
Size: 13 MB
Description:
The Census gives us a breakdown of the percentage of households by the Geographic Area (PUMA) that does not have an internet connection at home. This data is integral in mapping out which area of LA county is most in need. The latest data set is 2019, which fits right before the COVID lockdown. We can get more granular by bringing in other factors like socioeconomic status and ethnicity.



Data Cleaning

To use the data, we remove unnecessary and duplicate data sets to ensure we only observed what was relevant to our hypothesis. We first started by using Excel to quickly explore the CSV files and delete the unnecessary columns. Using this process, we were able to quickly remove/trim fields and create new custom fields. From Excel, we moved onto Jupyter notebooks to clean and merge the data further so we could leverage Python, Pandas, and Seaborn to complete the process-heavy cleaning.



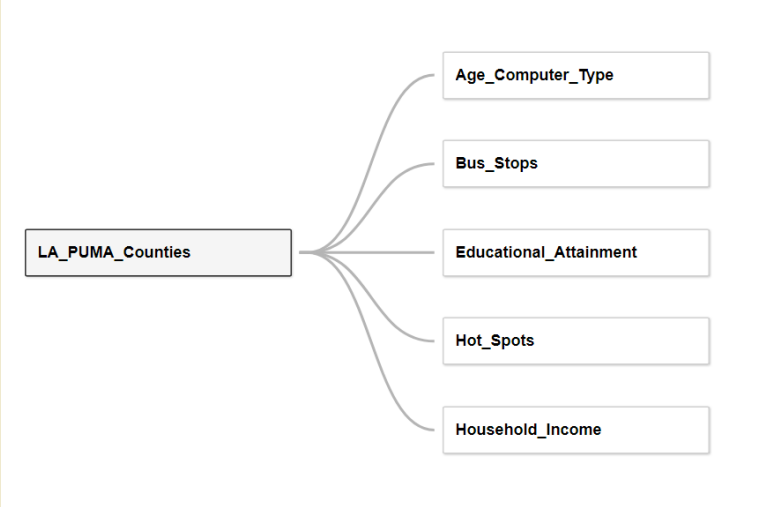
Using geospatial analysis, we were able to connect our various data sources by the GEOID. Leveraging Python libraries like shapely within a Jupyter notebook, we were able to analyze the PUMA shapefiles from the US Census and detect if the Wi-Fi or bus stop nodes were contained within them. This allowed us to create the visualizations in Tableau and our EDA.

```
match = []
for j, row in df.iterrows():
    p1 = Point(row['X'], row['Y'])
    found = False
    for index, feature in gj.iterrows():
        poly = shape(feature['geometry'])
        if (poly.contains(p1)):
            found = True
            match.append(feature['GEOID10'])
            break

    if (found == False):
        match.append("-")

df['GEOID10'] = match
```

One of the complexities we faced as we cleaned the data was attempting to link our zoning data. This required a lot of computational power to perform the nearest neighbor search between bus stop locations and zone. After letting the process run for 36 hours, the nearest zone was found, the zone was introduced as a new column within the Bus Stop data. We would like to use this in a future iteration of our modeling.

Data Engineering

Using the Census data which was cleaned previously, we created three separate tables featuring age, income, and education. This was used to create the income and education slides within our dashboard as well as providing analysis for our EDA. To filter and view our data by PUMA, we took the injected GEOID created previously into each table and linked them to the shapefiles in Tableau.



Custom sets were then introduced within our dashboard to include/exclude data based on certain criteria. One of the sets created featured PUMAs that included bus stops within them. This allowed us to decrease the number of errors when attempting to find the ideal PUMA for internet infrastructure.

Analysis

Through our analysis, we identified internet “dry spots” that -- if covered by Los Angeles County -- could reduce the internet resource gap that less-wealthy communities tended to experience, thereby helping to close the digital divide.

01. Unequal Distribution of Bus Stop in L.A.

Despite Los Angeles County currently containing a variety of libraries, cafés, and other places where people can find free Wi-Fi, many of its residents still live in areas that do not have Wi-Fi hotspots within walkable reach. One way to solve this would be to establish Wi-Fi hotspots at strategically chosen bus stops throughout the county.

An observation made with the bus stop data when viewed geographically on a map was that bus stop coverage per county was not equally distributed throughout the LA County regions. This proved to be a challenge when considering our problem statement and our intention of finding optimal places to place Wi-Fi internet nodes. This led us to hone into the viable counties which were the ones with the most coverage (Fig. 1).

We prioritized counties by available bus stops and density. Since South County, South Central, North/Unincorporated, and South West counties lacked available bus stop data--and therefore the infrastructure to support our goal--we removed them from consideration.

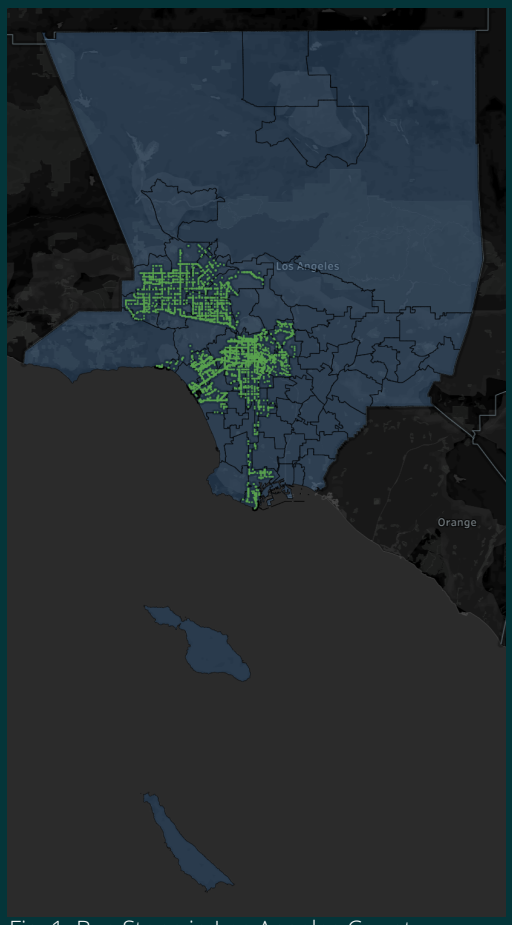
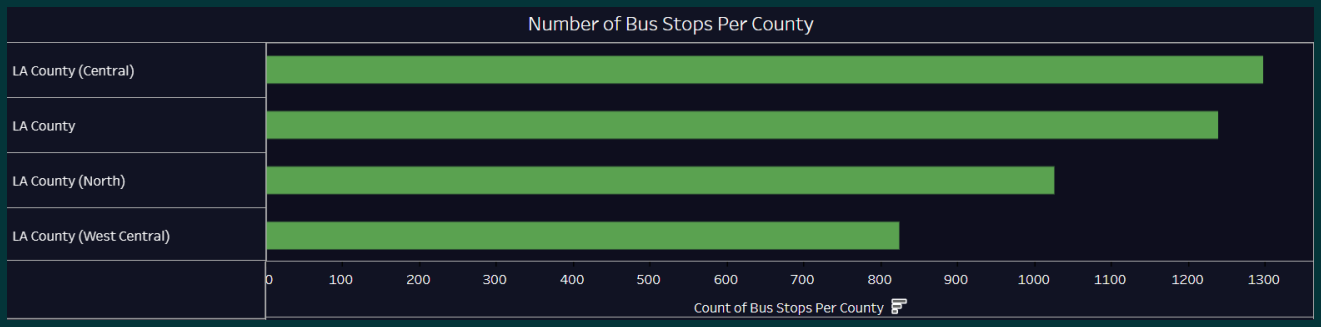
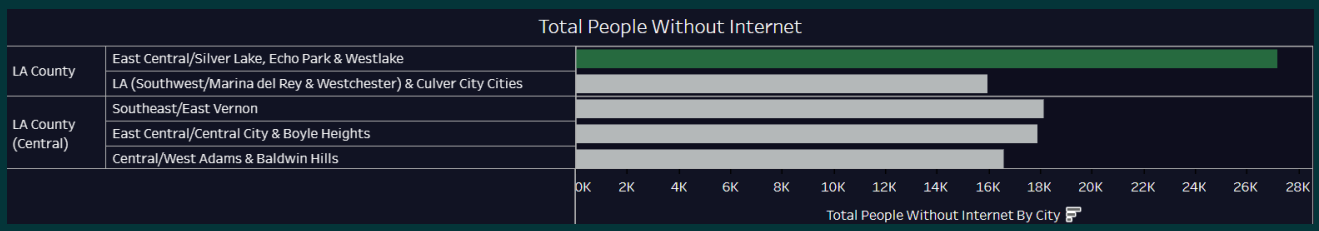


Fig. 1: Bus Stops in Los Angeles County



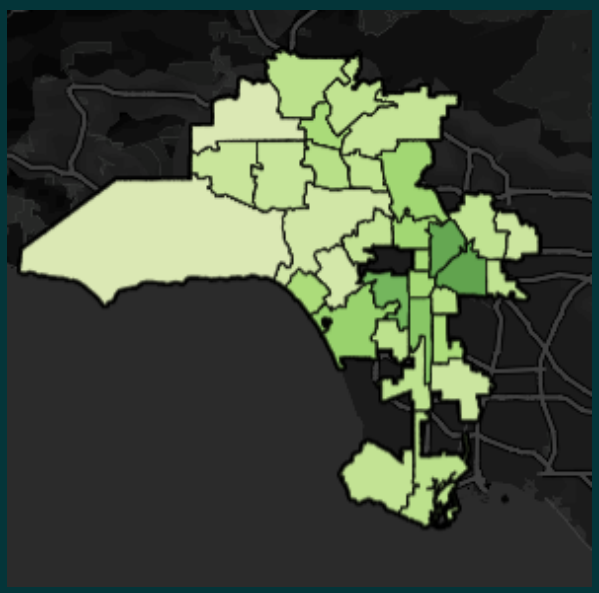
The chart above depicts the number of bus stops per county site. As we observe the chart, we see a majority of bus stops are within Los Angeles County and Los Angeles County (Central). Valley West contains approximately 25 times as many bus stops as South Bay. As we dive deeper, we look up internet disparities within the PUMAs within the individual counties.



Since our goal is to establish internet infrastructure in the area most impacted with low internet adoption, we determined that even though LA County (Central) had the most population, the most impacted PUMA was in LA County. This was the East Central/Silver Lake, Echo Park & West Lake PUMA.

02. Internet by Income Bracket

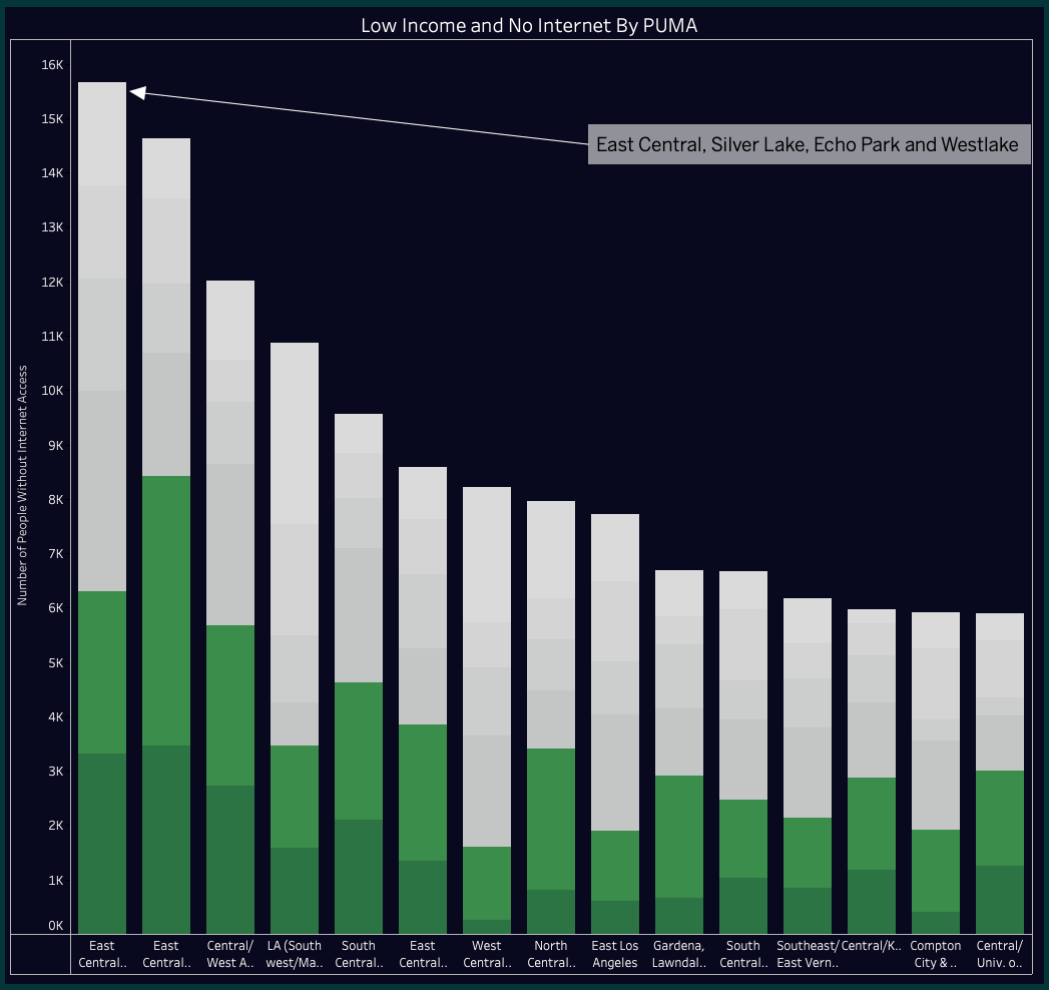
A trend appears between all counties in which a majority of income brackets between $0-$19,999 have the highest amount of percentage without internet. Southeast LA is the only county that does not follow this trend. Contrary to our initial thoughts, the income bracket within the $0-$20K range with the highest frequency of people without internet access was the income bracket $10,000-$19,000. Surprisingly, households with income below $10,000 had less overall people without internet access.



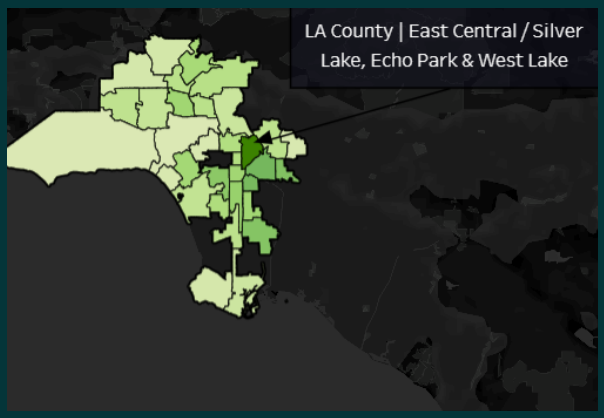
T-Test For Income < $20K Without Internet	
Central Vs. County Below	P-Value
Los Angeles County (Southwest)	0.39967367
Los Angeles County (North)	0.502173912
Los Angeles County (South Central)	0.547188748
Los Angeles County (Central)	0.665331654
Los Angeles County (Northwest)	0.950451957
Los Angeles County (South)	-
Los Angeles County (West Central)	-

At the 10% level of significance, there is not enough evidence to support the case that Los Angeles County has a higher proportion of people without internet access and making < $20K than any other county with bus stops. Thus, instead of focusing on the Los Angeles County with the most statistically significant proportion of people without internet access and making < $20K we focused on the city with the highest total number of people within this group.

Observing the above, Los Angeles County which is composed of East Central, Silver Lake, Echo Park and Westlake have the highest levels of Internet disparities by income levels compared to neighbooring counties.



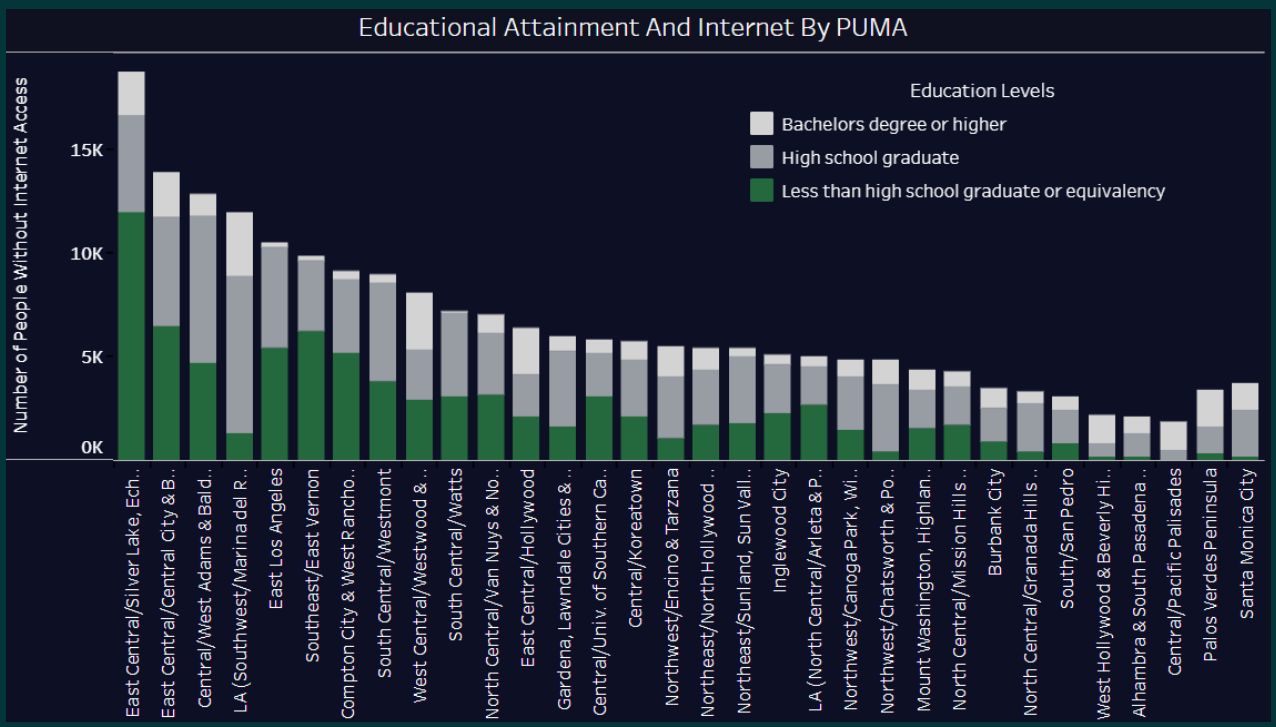
03. Educational Attainment and Internet



To select the best possible locations to place Wi-Fi hotspots, we explored the relationships between internet access and low education levels.

One of the most significant findings from the Education Attainment statistics was the high rate of people without a high school diploma in PUMAs where internet access was limited.

Analysis of the data showed over 50,000 residents in LA County (Central) and LA County combined with less than a high school education and lack of internet access. Directing our focus to these areas, we strategically filtered for the most accessible bus stop locations benefiting the community.



At the 10% level of significance, there is not enough evidence to support the case that Los Angeles County has a higher proportion of people without internet access and without a high school diploma than any other county with bus stops. Thus, instead of focusing on the Los Angeles County with the most statistically significant proportion of people without internet access and without a high school diploma, we focused on the city with the highest total number of people within this group.

T-Test For Less than High School Diploma	
Los Angeles Vs County	**P-Value**
Los Angeles County (Southwest)	0.189547998
Los Angeles County (North)	0.278290458
Los Angeles County (Northwest)	0.43020522
Los Angeles County (Central)	0.548507715
Los Angeles County (South Central)	0.699577401
Los Angeles County (South)	-
Los Angeles County (West Central)	-

Dashboard

To visualize our data interactively, we opted to use Tableau. This greatly facilitated our hosting and data analysis to one platform. It also simplified the technical feasibility for others in the group to help as no one had to another coding language.

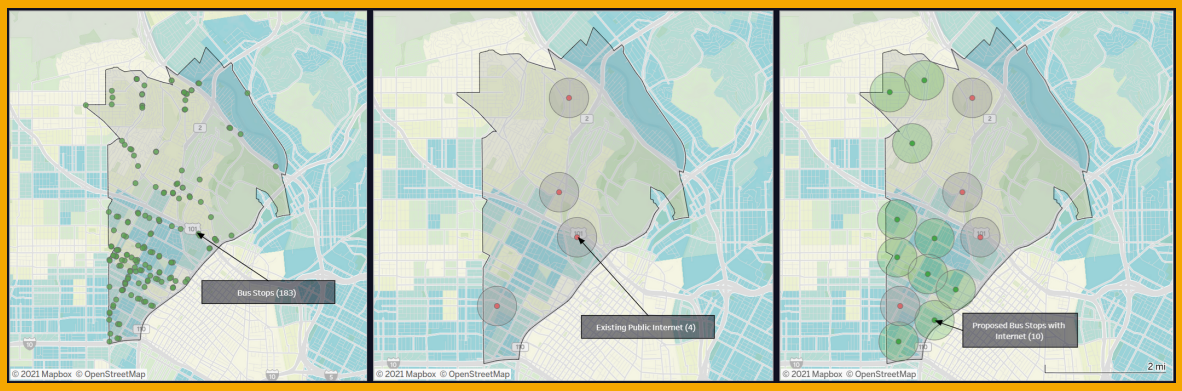
User Experience

As we navigated through our user experience options, our mentors mentioned checking Tableau's Viz of the Day. Once we understood the story mode capability of the tool, we translated our funneled analysis of our EDA into screens. This resembles a familiar approach to navigating a PowerPoint deck.

We started off by creating a mock-up of the product using Google slides. Once we understood the layout and storytelling we wanted to guide the user through, we began work on the dashboards. The five main dashboards created were the intro, bus stop locations, income, education, and conclusion.

Interactions

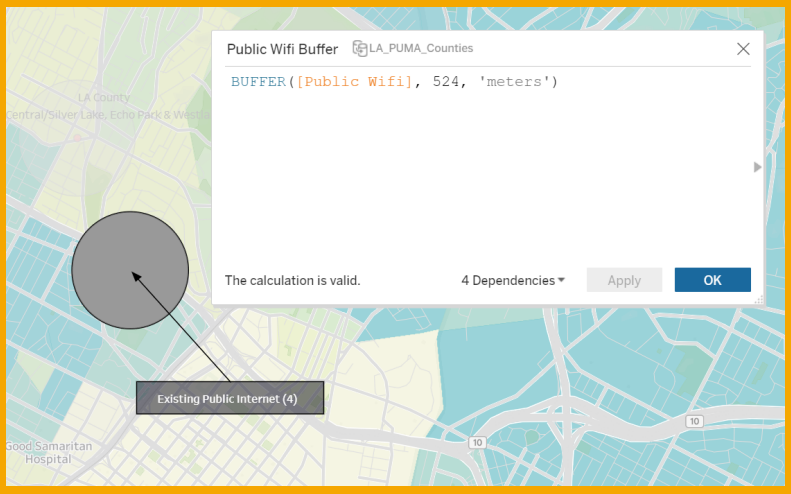
An important and large chunk of the storytelling we wanted to showcase was the interactivity between the end-user and our data. To accomplish this, we were able to leverage special filter functions for data to interact with other charts. This allows a user to click on charts and tables and have the map update on the currently selected state. This interaction serves as more independent exploration of the data and helps reinforce our conclusion.

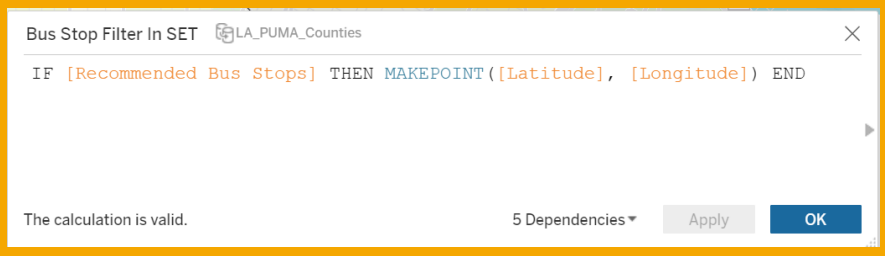


Custom Fields

There are some data points that are unique to our Tableau environment. These include special sets to calculated fields.

An example can be seen when using a radius buffer on the walkability of a potential bus stop. This is used for wifi and bus stop locations.





Filter were created to allow us to constrict the data present on the map. One example of this is the PUMAs that do not have infrastructure should not appear on the map.

Next Steps

Our team believes we can have a more impactful project by validating our assumptions and creating a better data model by collaborating with local governments.

01. Create a data model

Instead of selecting bus stops manually, we would use a maximal covering location problem model to determine optimal bus stop locations.

02. Validate asumptions

Assumptions were made based on previous studies about transit stops, which would need an additional study with this business case as conditions may have changed.

03. Complement data

We need more information about the data and conditions of current infrastructure like the bus stop type (bench, none, covered), electrical provision, etc.

04. Add filters

Leverage the bus stop zone type to prioritize areas of high residential or commercial populations, especially those with high population densities.

05. Collaboration

We would like to partner with the local government to prioritize the bus stop nodes on a weighted system of feasibility scale

Conclusion

Our team was able to determine the proper places where a local government as the L.A. County can use its existing infrastructure to provides Wi-Fi access point to those most in need of a broadband access.

Locations

Bus stops in **East Central, Silver Lake**, **Echo Park** and **West Lake** regions are the best county-owned places with a higher benefits & impact.

Great Impact

1080% increase in percentage of region with Wi-Fi hotspots on walkable reach

Benefits

Communities in need of internet connectivity would have a better chance to close the economic and educational gap between those who already have broadband access.

Organizations and persons in closer contact with local governments can validate and complement this report and analysis. Later on bootstrap initiatives to make this a reality.

Acknowledgements

A special thank you to all the mentors and the rest of the DS4A team for fostering a community of learning and data fluency.

Project Team

Francisco Razo Gomez

Espartaco Palma

Gerardo Mora

Jassmine Rabii

Marlene Plasencia

Mentor Team

Victoria Morgan

DS4A TEACHING ASSISTANT

Matt Gardner

DS4A PROFESSIONAL MENTOR

Sia Seko

DS4A TEACHING ASSISTANT

Liam Hanham

DS4A PROFESSIONAL MENTOR

References

Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. 2021. http://doi.org/10.18128/D050.V16.0

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant43 Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps 44 using variable service areas. Transportation, 41(1), 193-210.

Thompson, G. (2020, November 30). Two thirds of the world's school-age children have no internet access at home, new UNICEF-ITU report says. Retrieved May 01, 2021, from https://www.unicef.org/press-releases/two-thirds-worlds-school-age-children-have-no-internet-access-home-new-unicef-itu

Bridging the Digital Divide. (n.d.). Retrieved May 01, 2021, from https://www.fcc.gov/about-fcc/fcc-initiatives/bridging-digital-divide-all-americansAdditional Data Sources