

Dataset Description and Definitions:

Census data 2020 is not currently available. In order to look at block group level data, we have analyzed American Community Survey 5 year data. There is data available for 1 year estimates, but only at the county level. We have looked at variables related to internet access, including the proportion of residents with no internet access and the proportion with broadband internet subscriptions. We compared this with the racial breakdown and mean income of the residents. We were interested in comparing this with other variables such as education status, but not all variables are available at the block group level.

Limitations:

Our main limitation has been that in order to get block group level data (the closest approximation to neighborhood level we have), we have to look at 5-year aggregated data, and not all variables are available. There are other datasets related to internet speeds that would be useful to tie in, but getting data for a small geographic area and tying it back to the units the Census uses (in this case census block group), is challenging.

We are also mainly working with survey data, with all the limitations that come with them, including people who do not respond or do not respond accurately.

If the 2020 Census data gets released in time and includes the broadband variable, we can then look at census block level data. Otherwise, we will have to stick with census block group level 5 year American Community Survey estimates.

Preliminary Conclusions

When it comes to those who have no internet access at all, neighborhoods that have lower income, more African American residents, and more Native American residents are more likely to have a higher proportion of residents with no internet access at all. These trends are the same for those who lack access to broadband subscriptions.

Project Scope

No change.

Goals for final report following meeting with client

```
[9] data["% no internet"] = data["% no internet"].apply(lambda x: 0 if x <= 10 else 1 if x
[10] dataset = data.drop(columns = ['% no internet', '% internet with broadband subscription
labels = data["% no internet"].values.astype(np.float32)

logistic_regression_classify_mnist(dataset, labels)

Using 16bit native Automatic Mixed Precision (AMP)
GPU available: True, used: True
TPU available: False, using: 0 TPU cores
IPU available: False, using: 0 IPUs
/usr/local/lib/python3.7/dist-packages/pytorch_lightning/trainer/trainer.py:733: Lightn
"trainer.fit(train_dataloader)" is deprecated in v1.4 and will be removed in v1.6."
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]

| Name | Type | Params
-----|-----|-----
0 | layer_1 | Linear | 60
-----|-----|-----
60 | Trainable params
0 | Non-trainable params
60 | Total params
0.000 | Total estimated model params size (MB)
/usr/local/lib/python3.7/dist-packages/pytorch_lightning/trainer/data_loading.py:111: U
f"The dataloader, {name}, does not have many workers which may be a bottleneck."
/usr/local/lib/python3.7/dist-packages/pytorch_lightning/trainer/data_loading.py:407: U
f"The number of training samples ({self.num_training_batches}) is smaller than the lo
Epoch 299: 100%
(89.86350297927856, 88.67543339729309)
```

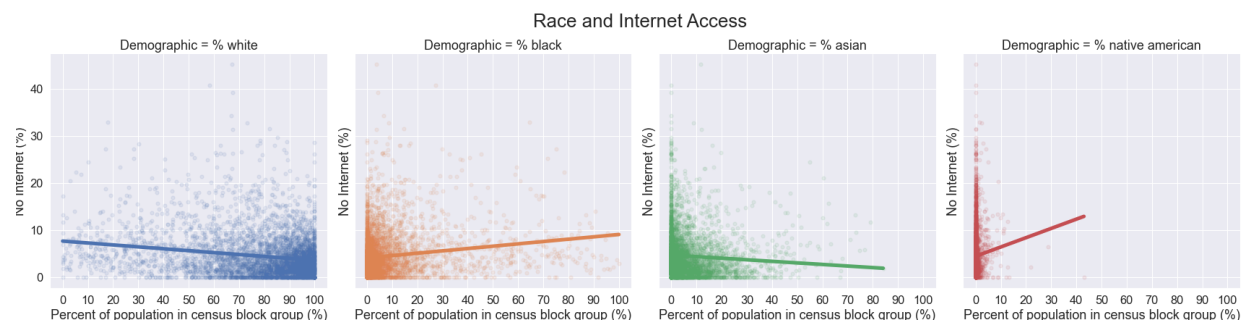
Following our client meeting on 11/17 we have a clearer understanding of the desired end product. We currently have heat maps of Massachusetts counties showing the percentage of people without internet access and with broadband access. Our client asked if we could find a way to make the maps more intuitive. Our goal now is therefore to plot the heat maps on top of a regular map, such that it becomes easier to relate census block groups to real areas of the county.

Likewise, we have tables showcasing the demographic setup of the areas with the lowest access to the internet. The client again asked if we had a way to translate those table rows into geographic locations. We therefore also want our table to include a zip code, so that it is easier to locate where in the state the people with reduced internet access live. We have found an API that can make this conversion from census tract however, it will not be part of this deliverable.

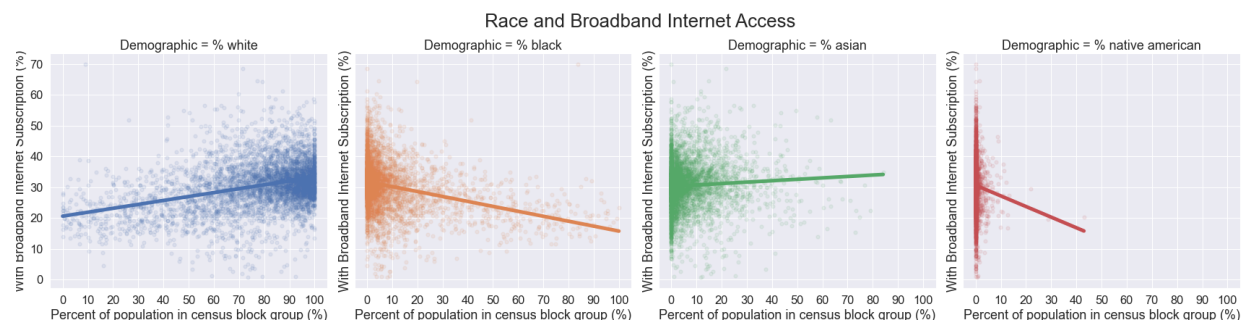
Finally, as seen in the picture above, we have built our first iteration of a logistic regression model which takes as input variables the demographic composition of the respective block group as well as its mean income level, and predicts percentage of internet access as discretized into ranges of 10 percentage points from 0 to 100. This model performs with 88% test accuracy, indicating that anything more than that would require a model that would allow for datasets that are not linearly separable. Hence, we may choose any number of methods that would be appropriate given these constraints, i.e. SVM or random forests, while still allowing our results to be interpretable. We will also create synthetic examples that emphasize particular features while nulling the rest, to create intuition about how the model associates individual features such as demographic composition of a block group to predict internet access. Our next steps will involve implementing as much, notwithstanding the extension of this analysis via other models as indicated above.

Figures and Tables:

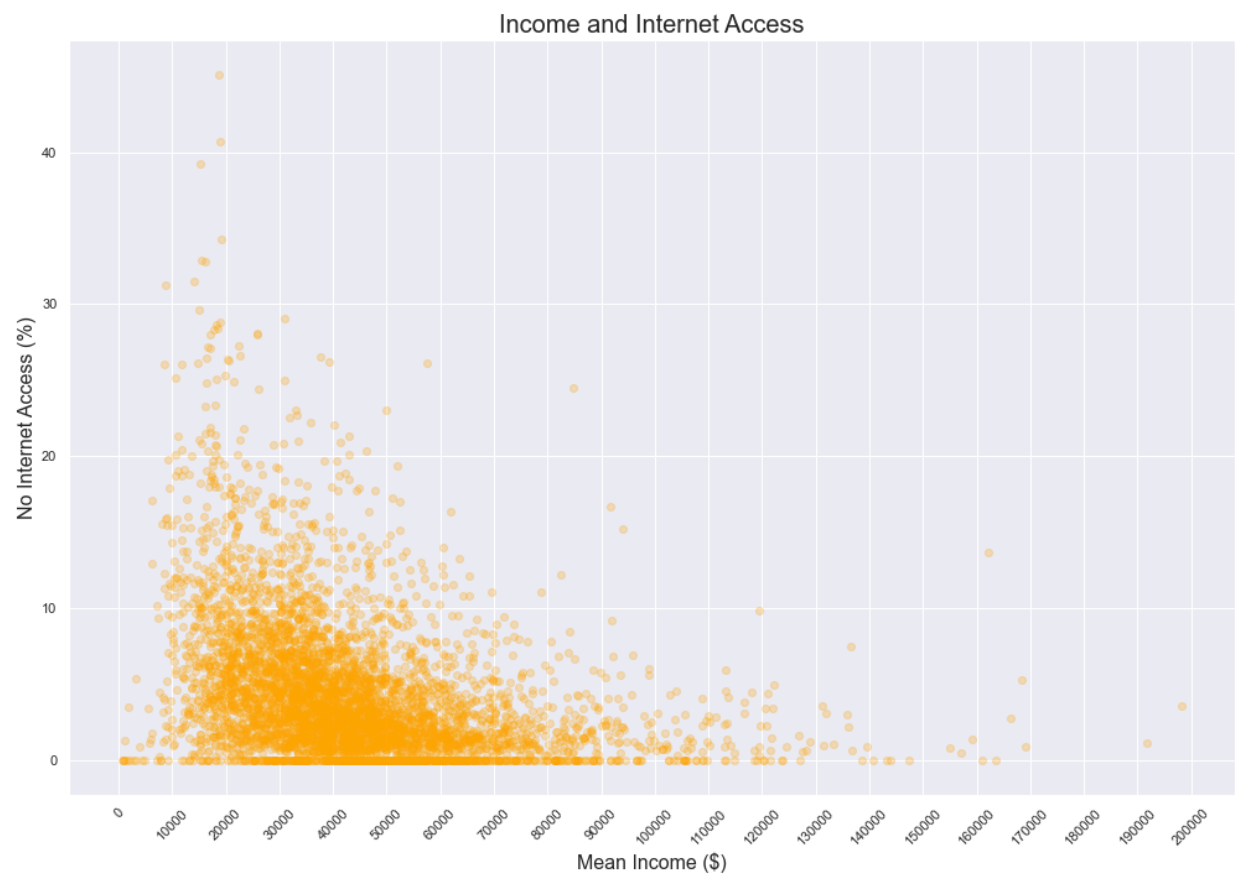
Racial breakdown of block groups, as compared with percent without internet access:



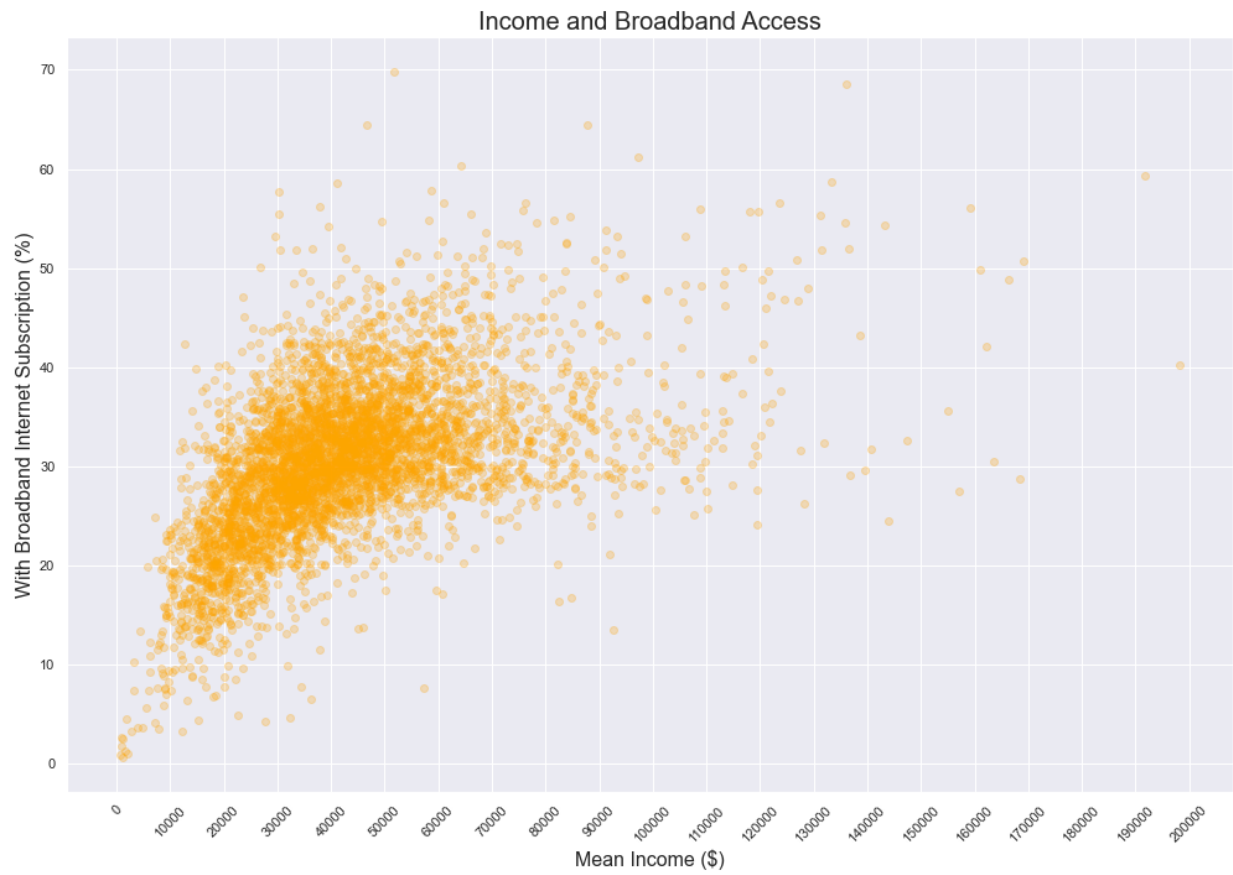
Racial breakdown of block groups, as compared with percent with a broadband subscription:



Income and internet access:

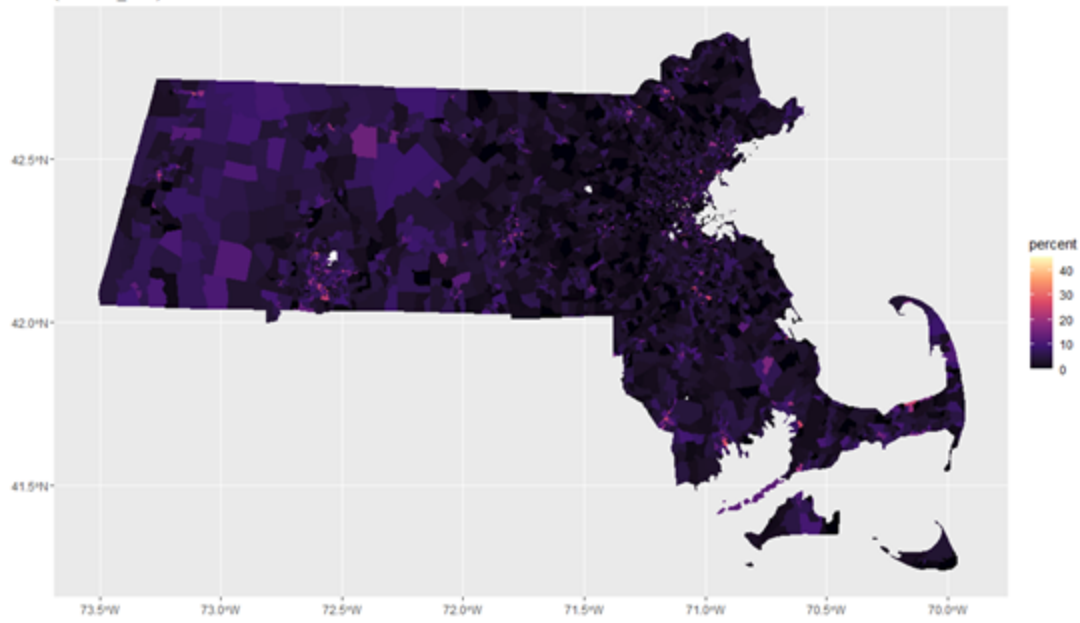


Income and broadband internet subscriptions:

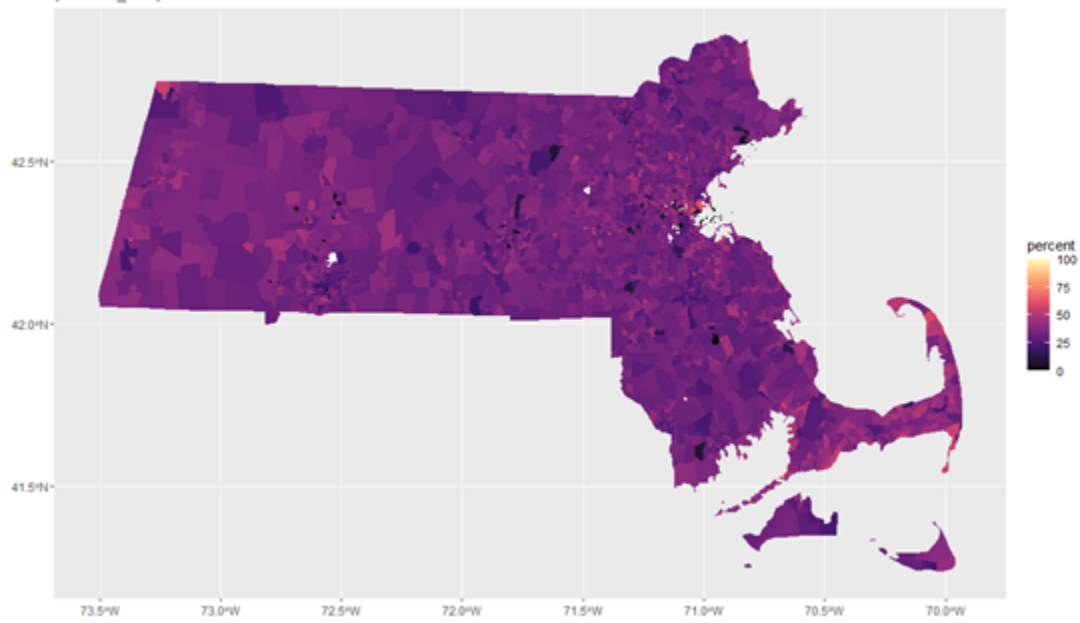


Massachusetts:

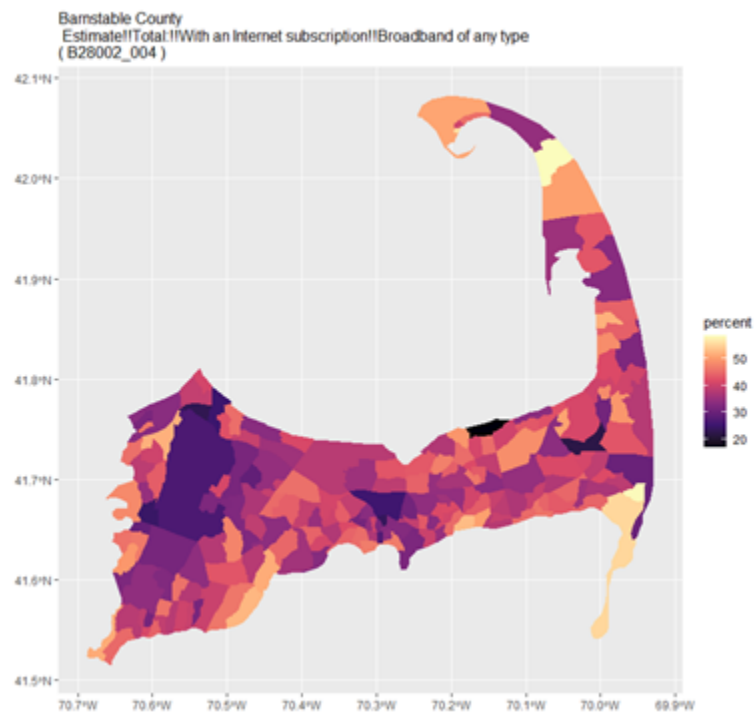
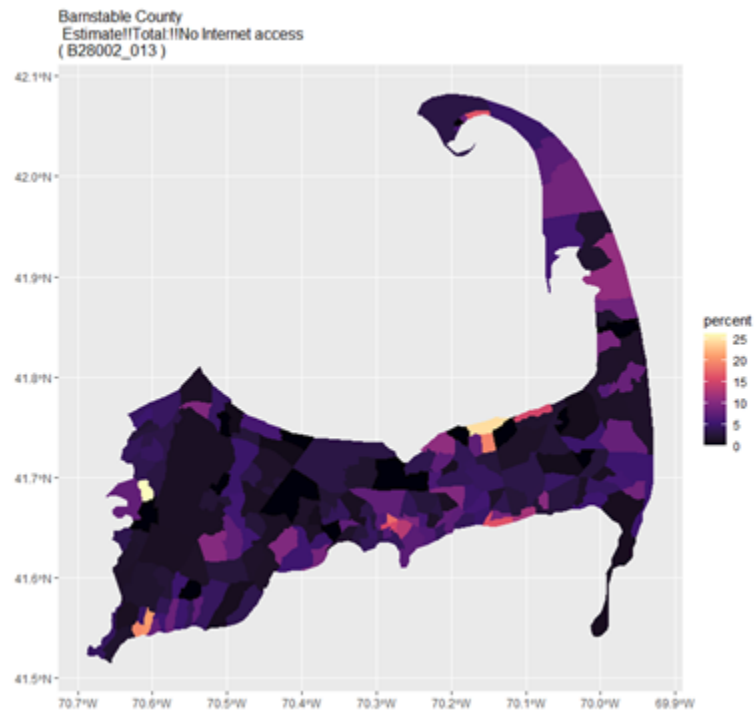
Massachusetts
Estimate!!Total!!No Internet access
(B28002_013)



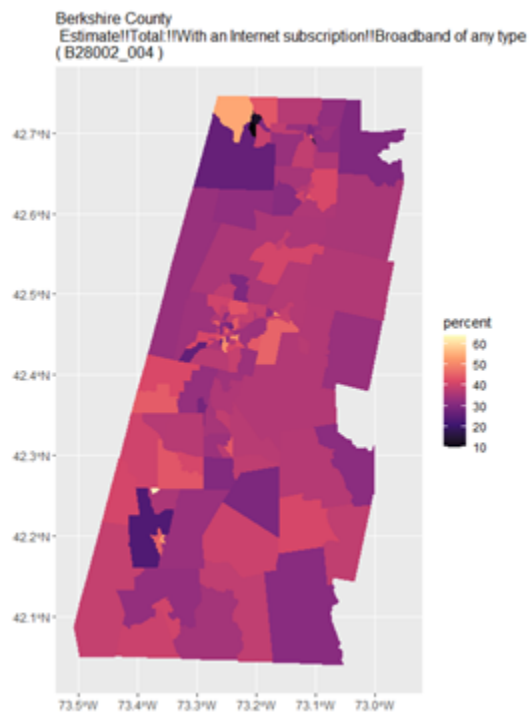
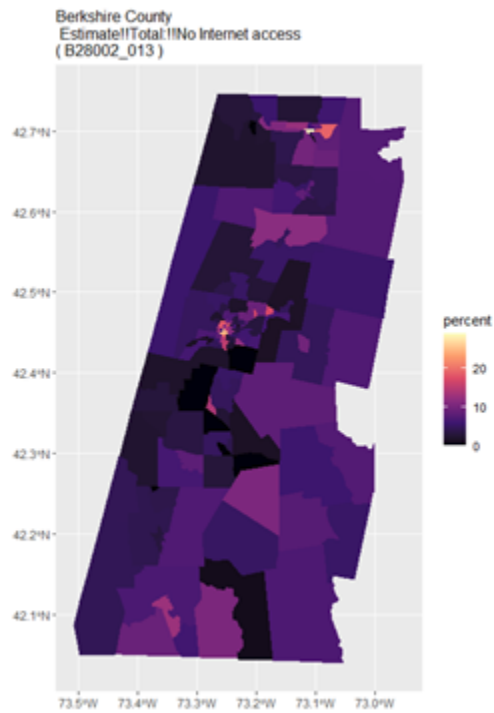
Massachusetts
Estimate!!Total!!With an Internet subscription!!Broadband of any type
(B28002_004)



Barnstable County:



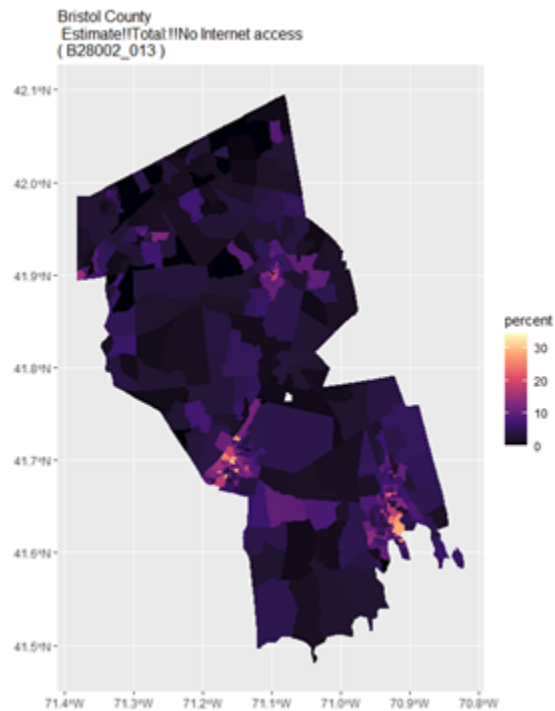
Berkshire County:

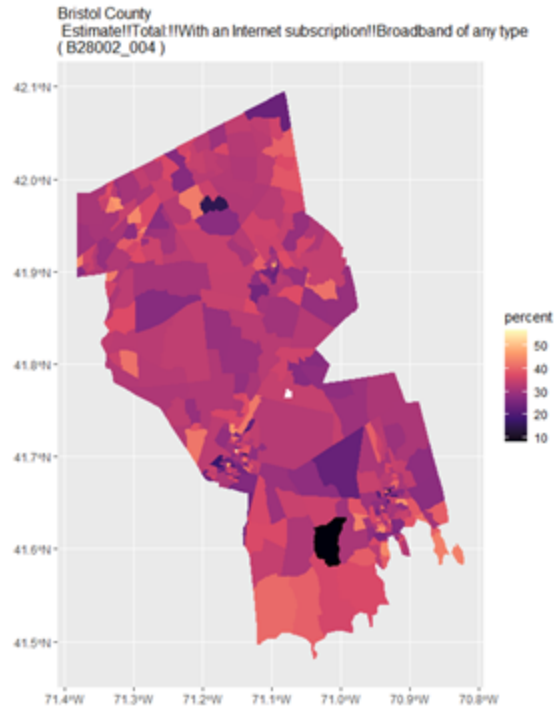


Berkshire County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
4	9353	Block Group 4	28.663793	28.232759	100.0	0.0	0.0	0.0	18301.0
2	9001	Block Group 2	24.905660	29.056604	89.2	1.8	1.3	0.0	21432.0
3	9001	Block Group 3	18.911175	46.131805	83.4	4.6	0.0	0.0	42373.0
1	9213	Block Group 1	18.843120	27.519720	89.8	0.0	2.8	0.0	26829.0
6	9002	Block Group 6	18.271120	29.076621	76.4	13.0	4.3	0.0	22537.0
3	9006	Block Group 3	17.995910	20.449898	67.5	27.6	0.0	0.0	16951.0
1	9011	Block Group 1	17.173913	31.521739	100.0	0.0	0.0	0.0	30468.0
4	9121	Block Group 4	16.932907	28.913738	97.6	1.1	0.0	0.0	24620.0
5	9009	Block Group 5	15.221239	51.504425	97.5	2.5	0.0	0.0	93983.0
1	9001	Block Group 1	14.226519	39.779006	76.8	12.3	1.5	0.0	21339.0

Bristol County:

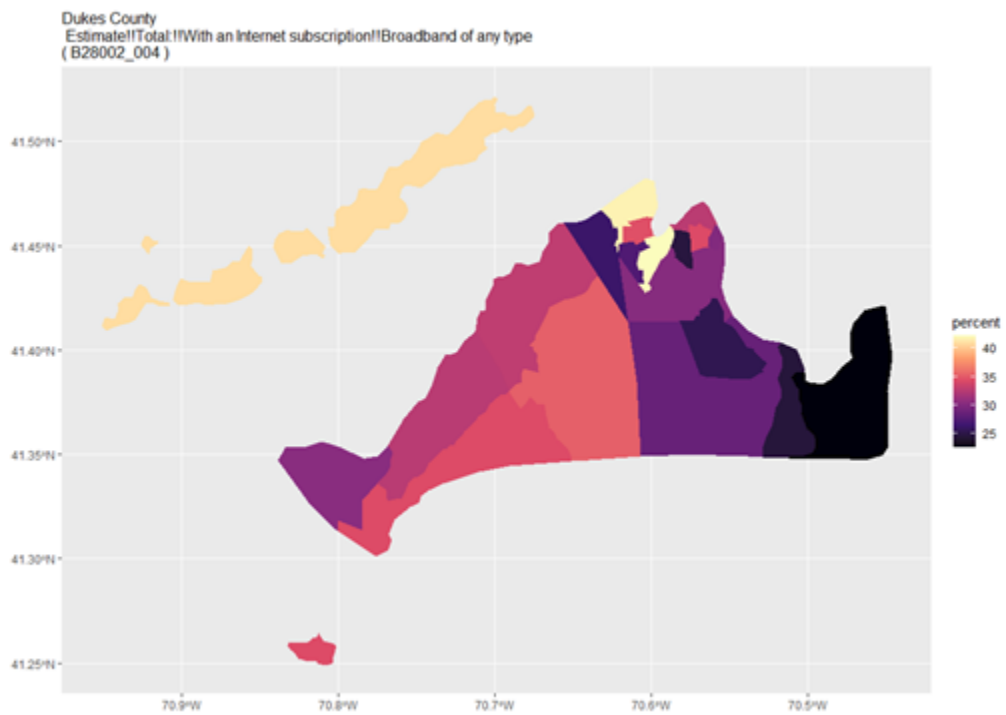
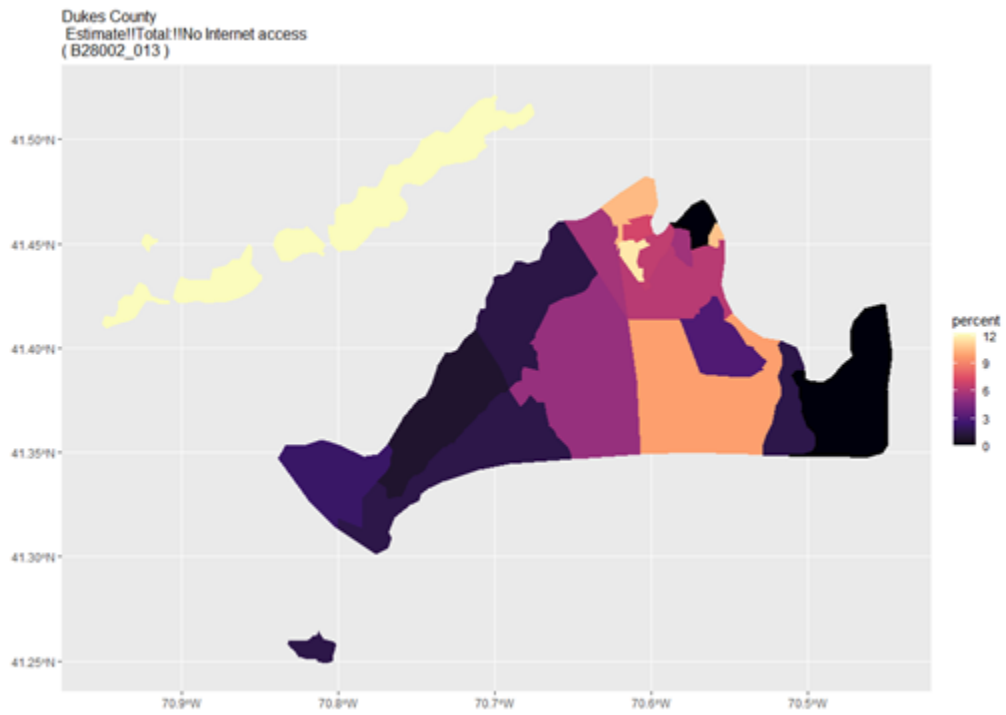




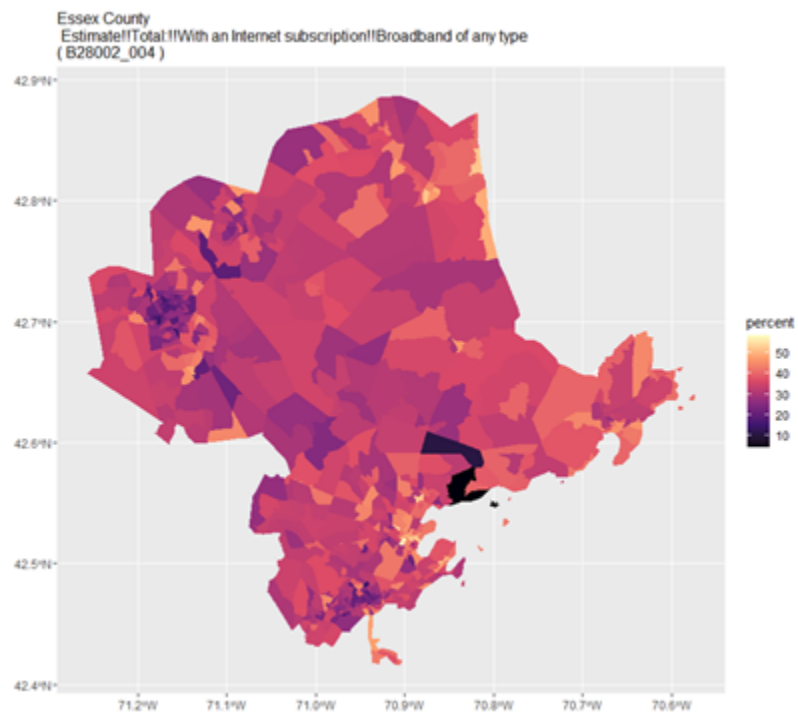
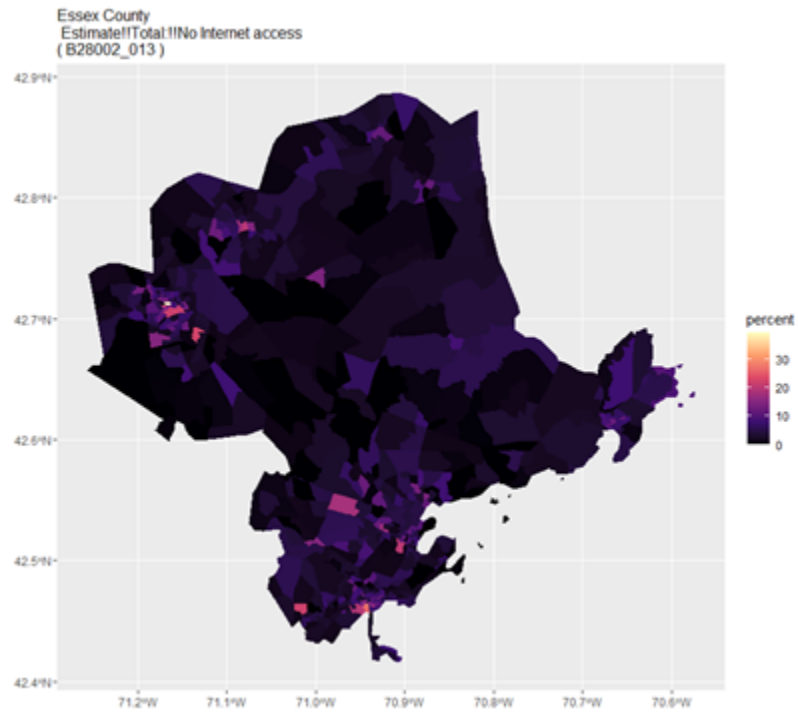
Bristol County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
2	6411.01	Block Group 2	34.251291	18.760757	66.8	6.5	0.0	0.0	19228.0
2	6402	Block Group 2	29.629630	28.104575	85.4	8.9	0.0	0.0	15034.0
1	6519	Block Group 1	28.322440	26.579521	30.3	8.1	0.0	0.0	17781.0
1	6406	Block Group 1	28.067227	24.873950	84.2	12.8	0.0	0.0	25936.0
1	6411.01	Block Group 1	27.979798	23.535354	77.1	13.3	0.0	0.0	17063.0
2	6421	Block Group 2	26.641651	15.103189	71.0	1.9	0.0	0.0	22577.0
1	6404	Block Group 1	26.509186	24.934383	88.7	7.3	0.0	0.0	37589.0
2	6527	Block Group 2	26.339969	28.483920	67.5	6.1	0.0	0.0	20233.0
2	6518	Block Group 2	26.284585	31.027668	62.5	5.9	0.0	0.0	20604.0
2	6506	Block Group 2	26.020408	12.500000	61.1	6.6	0.0	0.0	11765.0

Dukes County:



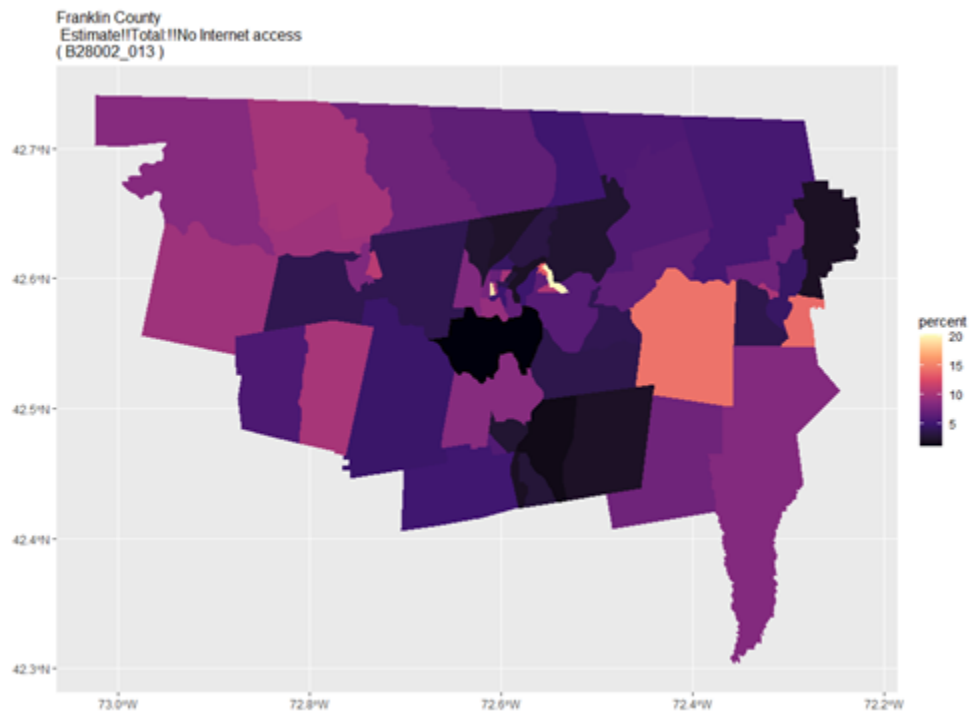
Essex County:

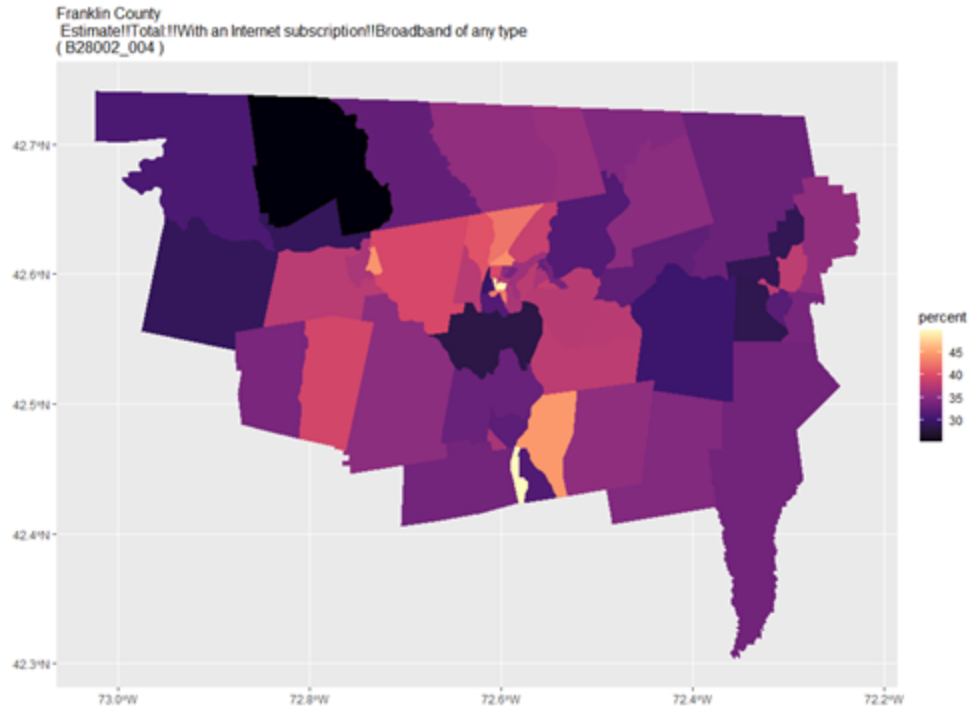


Essex County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
1	2511	Block Group 1	39.262821	13.621795	67.5	4.2	0.0	0.0	15258.0
1	2069	Block Group 1	28.383706	27.201051	78.3	17.1	0.0	2.9	18466.0
3	2069	Block Group 3	26.109391	39.834881	64.0	32.6	0.0	0.0	14791.0
3	2532.02	Block Group 3	22.548121	25.022915	90.0	1.3	5.3	0.0	31953.0
3	2082	Block Group 3	22.056921	34.282018	88.4	4.6	5.3	0.0	40126.0
1	2501	Block Group 1	21.835116	22.869955	49.7	6.1	0.0	0.6	23369.0
2	2511	Block Group 2	20.703125	15.364583	42.1	15.1	0.0	0.0	18134.0
2	2047.01	Block Group 2	20.146765	20.613742	58.8	13.1	6.5	0.0	22139.0
2	2601	Block Group 2	20.098039	19.747899	79.3	11.1	2.2	0.0	18110.0
1	2070	Block Group 1	19.757174	24.613687	47.0	27.6	0.0	0.0	18697.0

Franklin County

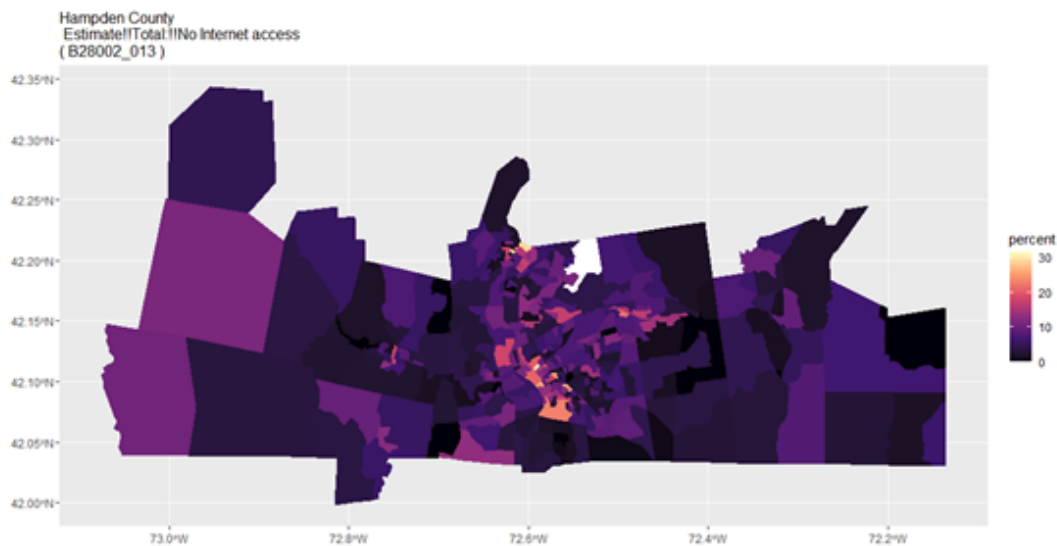


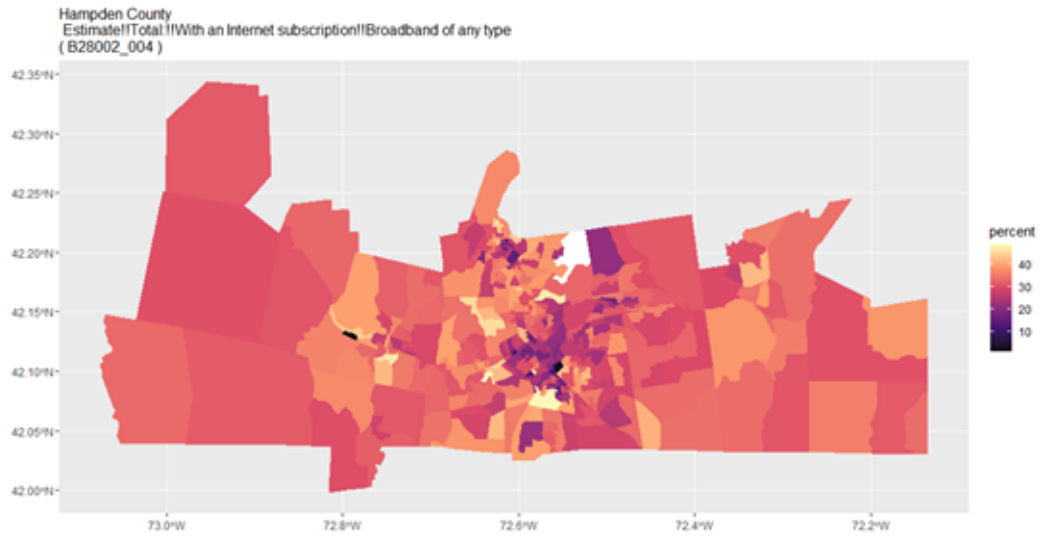


Franklin County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
4	407.01	Block Group 4	20.104439	32.114883	97.9	0.0	0.0	0.0	42906.0
1	413	Block Group 1	19.429199	29.198683	100.0	0.0	0.0	0.0	19564.0
3	404	Block Group 3	14.269142	4.640371	89.0	4.5	0.8	0.0	32244.0
3	405.02	Block Group 3	13.894888	31.605472	95.3	1.2	0.0	0.0	24714.0
2	407.01	Block Group 2	13.256007	30.985915	82.6	8.2	0.0	0.0	18634.0
4	415.02	Block Group 4	10.814249	39.058524	92.7	1.3	6.0	0.0	43130.0
1	407.02	Block Group 1	10.314875	35.287731	88.8	0.0	0.0	11.2	30777.0
3	405.01	Block Group 3	10.092961	34.528552	93.8	0.5	0.0	3.9	32651.0
1	415.01	Block Group 1	9.986684	23.701731	94.7	0.0	0.0	0.3	52493.0
2	412	Block Group 2	9.828010	27.395577	99.3	0.7	0.0	0.0	35712.0

Hampden County:

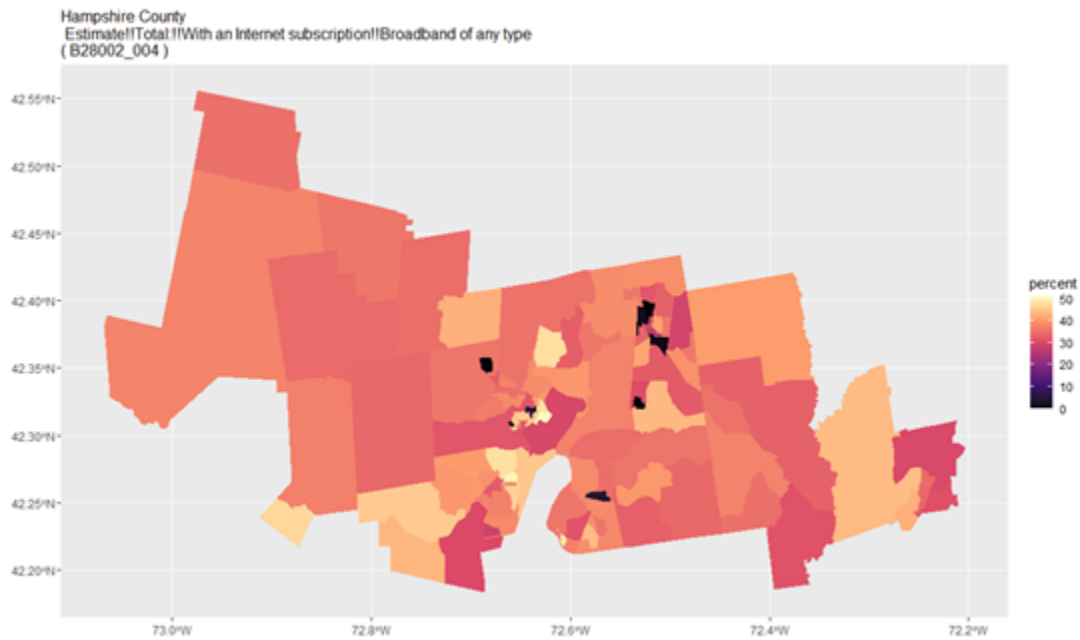
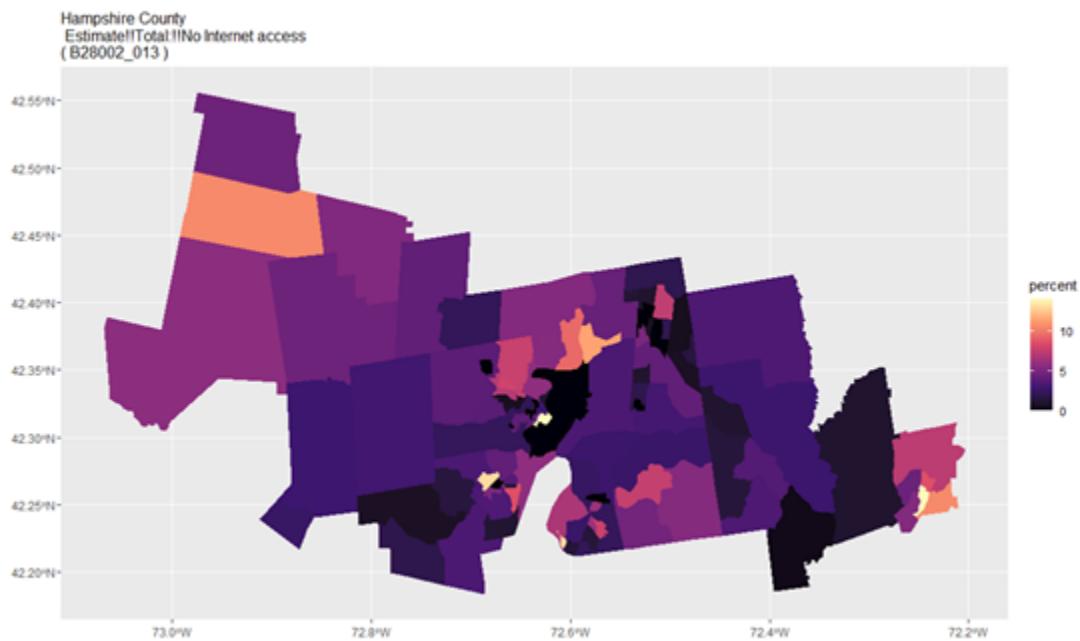




Hampden County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
2	8009	Block Group 2	31.524548	13.608958	82.0	14.6	0.0	0.0	14009.0
1	8117	Block Group 1	29.039813	29.274005	78.5	0.0	0.0	0.0	31003.0
1	8120.01	Block Group 1	28.821811	14.410906	89.0	3.5	0.0	0.0	18935.0
1	8023	Block Group 1	27.086882	13.628620	87.9	8.2	0.0	0.0	17075.0
2	8011.02	Block Group 2	26.027397	9.132420	79.2	8.2	0.0	0.0	8588.0
4	8018	Block Group 4	25.323475	17.929760	52.3	40.7	0.0	0.0	19930.0
5	8021	Block Group 5	23.048327	48.513011	67.3	1.1	13.4	0.0	33032.0
3	8001.02	Block Group 3	21.592775	13.464696	65.9	29.1	0.0	0.0	17045.0
2	8118	Block Group 2	21.508380	9.683426	97.1	1.3	0.0	0.0	16067.0
1	8011.01	Block Group 1	20.760234	25.087719	50.1	15.0	0.8	0.5	18058.0

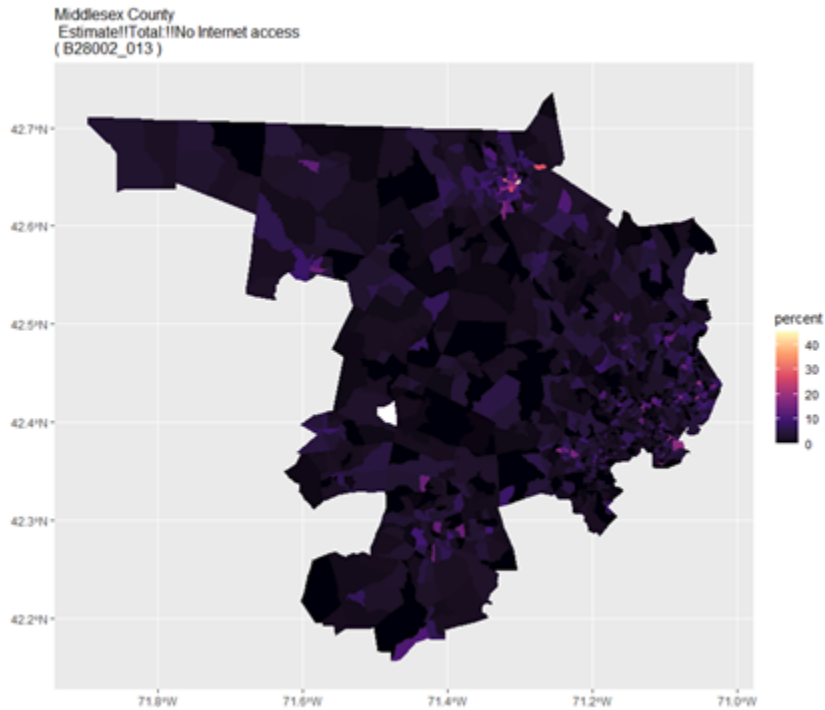
Hampshire County:



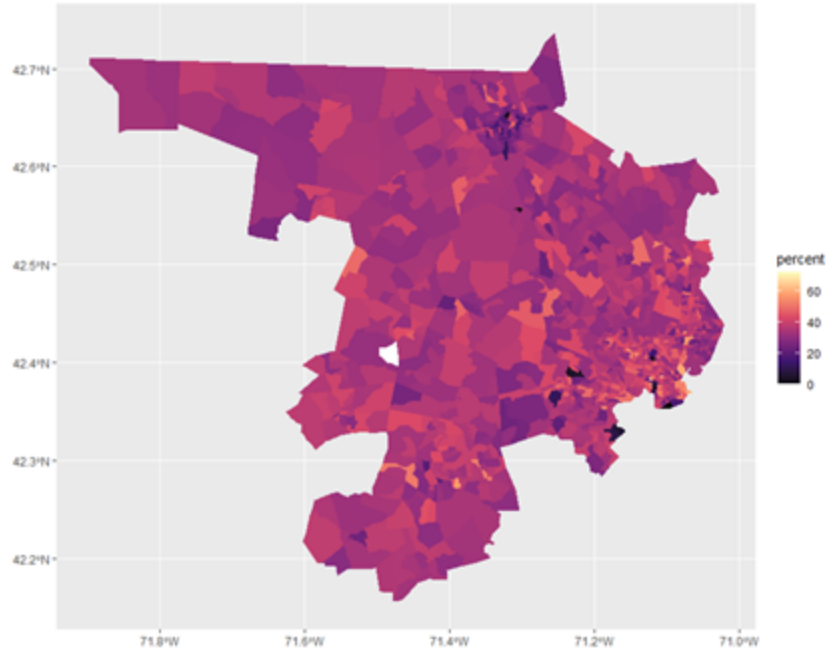
Hampshire County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
3	8219.03	Block Group 3	13.998704	48.282566	91.4	3.8	1.0	0.0	40839.0
2	8201.02	Block Group 2	13.976165	31.419285	100.0	0.0	0.0	0.0	24152.0
4	8224.02	Block Group 4	13.007160	36.276850	95.7	1.2	0.0	0.0	56401.0
5	8211	Block Group 5	12.997904	45.492662	99.8	0.2	0.0	0.0	38740.0
2	8214	Block Group 2	11.372064	33.745365	92.6	2.7	0.0	0.0	44212.0
1	8201.02	Block Group 1	10.570236	29.902643	93.5	4.6	0.0	0.0	33279.0
2	8227	Block Group 2	10.532544	26.627219	98.1	0.0	0.6	0.0	34204.0
1	8215	Block Group 1	9.371884	28.813559	96.5	0.4	0.0	2.2	33703.0
3	8201.02	Block Group 3	8.559201	27.627199	93.6	2.6	0.0	0.0	21770.0
2	8223	Block Group 2	8.547889	41.400618	92.8	6.0	0.0	0.0	32156.0

Middlesex County



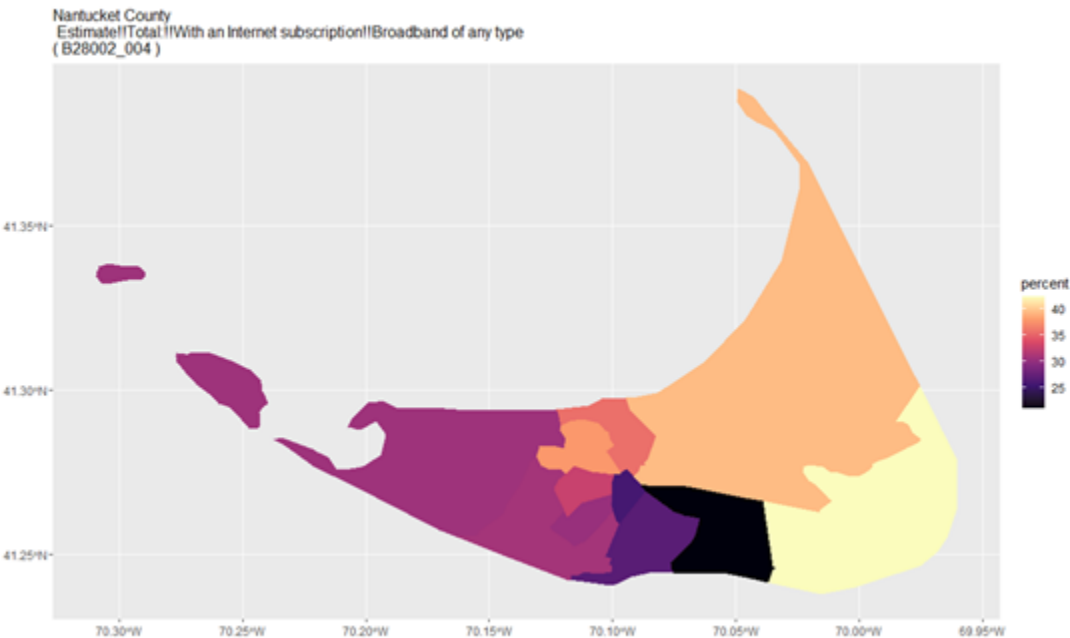
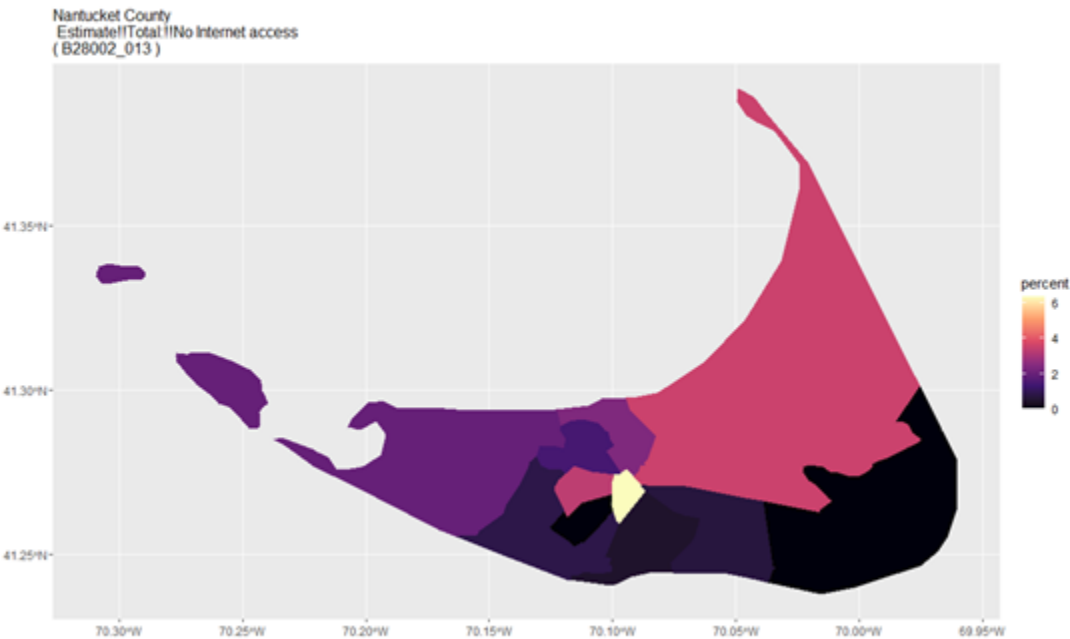
Middlesex County
 Estimate!!Total!!With an Internet subscription!!Broadband of any type
 (B28002_004)



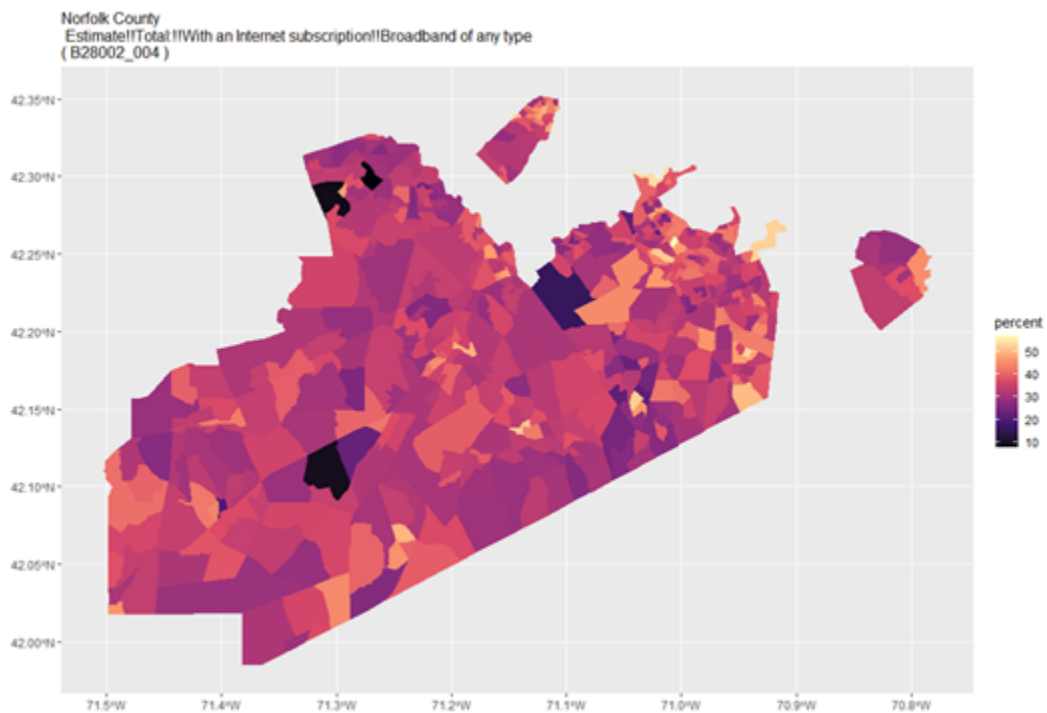
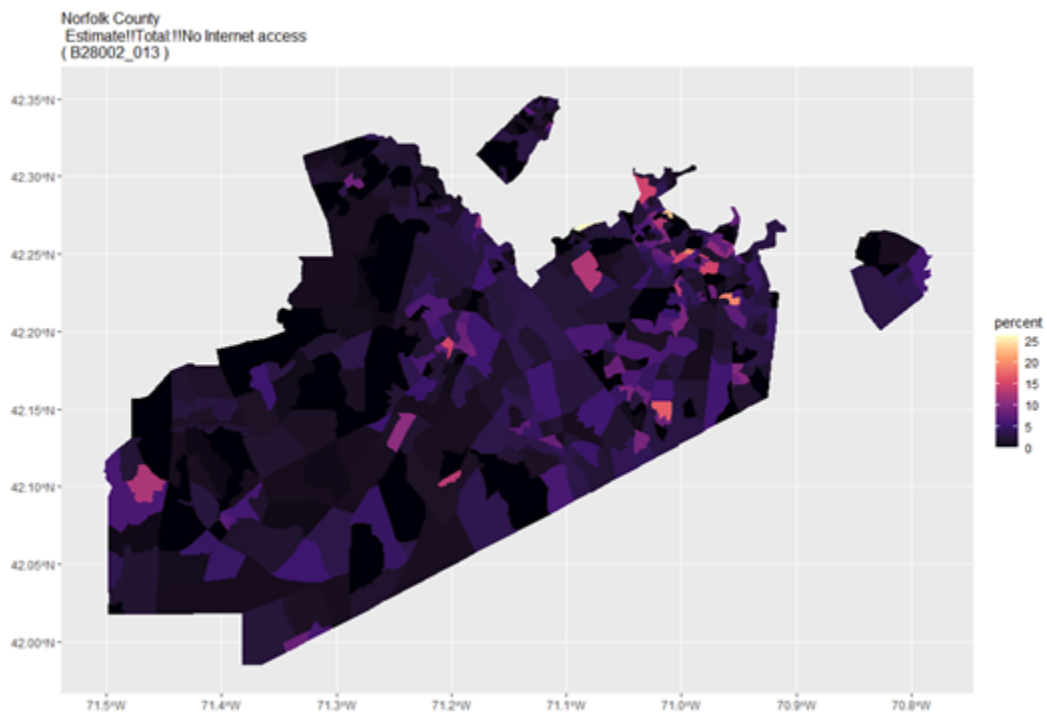
Middlesex County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
1	3124	Block Group 1	45.109780	14.171657	67.3	3.6	11.8	0.0	18630.0
1	3883	Block Group 1	31.295201	15.910585	67.6	3.7	12.0	0.0	8671.0
2	3143.01	Block Group 2	28.000000	18.400000	62.7	16.6	16.3	0.0	25954.0
2	3119	Block Group 2	26.457399	24.065770	60.7	6.4	19.7	2.2	16359.0
1	3104	Block Group 1	20.833333	12.500000	80.3	9.7	9.9	0.0	15540.0
1	3101	Block Group 1	19.166057	10.460863	63.4	20.6	7.5	0.6	12294.0
1	3507	Block Group 1	18.755981	30.526316	84.9	4.5	1.7	0.0	41172.0
1	3515	Block Group 1	18.452381	41.468254	60.5	4.2	17.5	0.0	42984.0
6	3418	Block Group 6	18.386243	31.746032	77.9	6.1	16.0	0.0	17035.0
1	3112	Block Group 1	17.984361	20.590791	37.4	14.2	38.8	1.0	18603.0

Nantucket County:



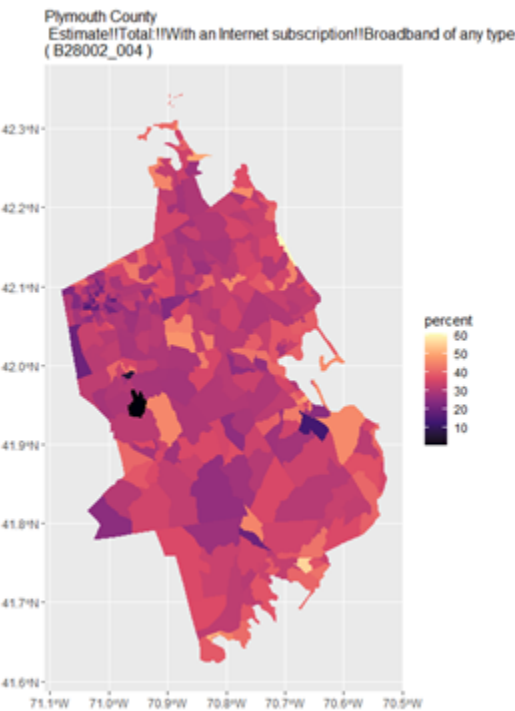
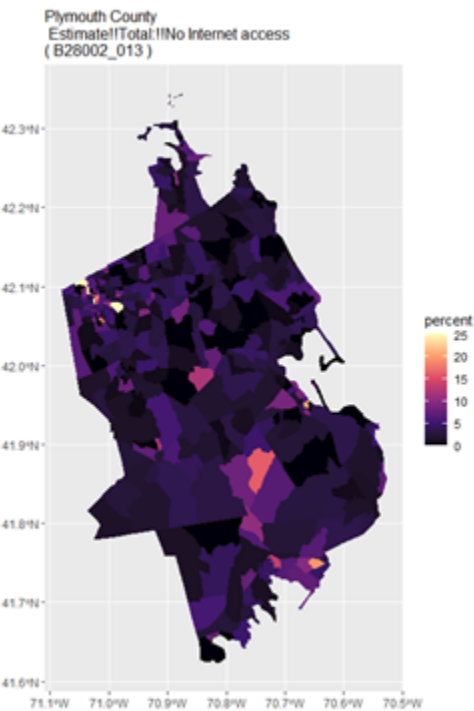
Norfolk County:



Norfolk County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
5	4163	Block Group 5	26.162019	26.162019	51.1	45.2	0.0	0.0	57567.0
1	4176.01	Block Group 1	23.044397	31.078224	83.7	3.2	13.1	0.0	49955.0
2	4177.01	Block Group 2	19.498290	35.689852	70.6	2.1	27.4	0.0	26364.0
1	4224	Block Group 1	19.281437	27.425150	91.3	0.0	1.8	0.0	29345.0
1	4211	Block Group 1	16.845329	24.119449	72.7	12.4	4.2	0.0	29084.0
1	4134.01	Block Group 1	15.091864	37.007874	88.1	11.8	0.0	0.0	38746.0
2	4179.01	Block Group 2	14.900662	33.869442	60.4	5.5	21.6	0.0	33955.0
4	4177.01	Block Group 4	14.769031	36.109304	63.3	4.5	29.3	0.2	33132.0
1	4173	Block Group 1	14.739884	37.861272	77.7	4.2	18.0	0.0	45351.0
4	4179.01	Block Group 4	13.728549	35.881435	45.7	0.3	45.1	0.0	26630.0

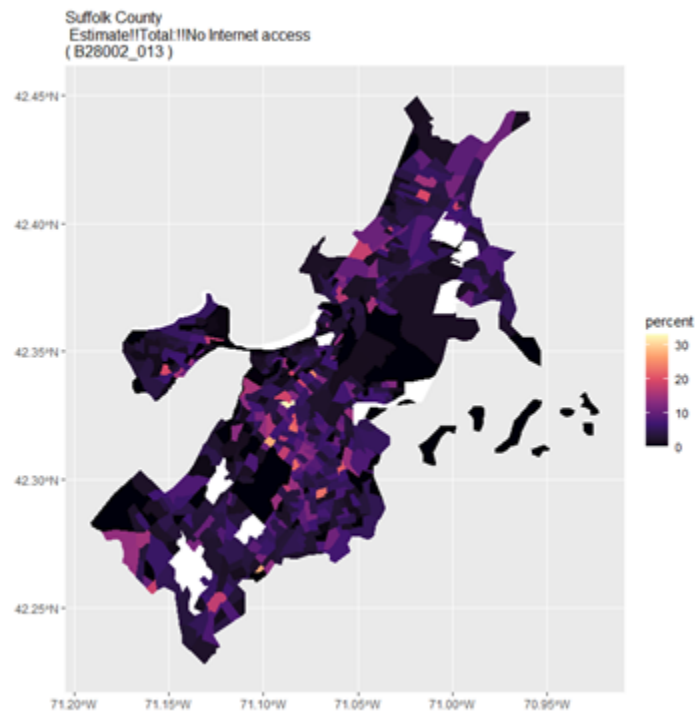
Plymouth County:

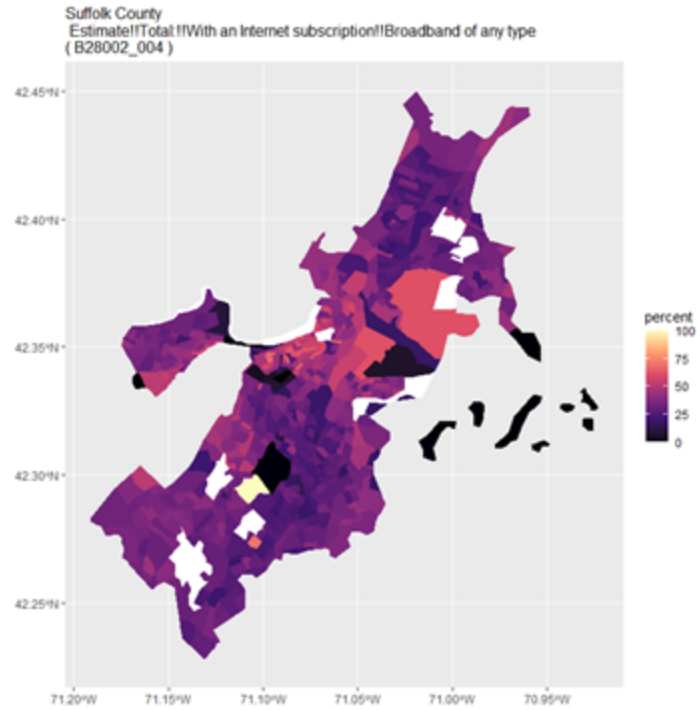


Plymouth County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
3	5108	Block Group 3	25.173210	16.974596	28.3	71.7	0.0	0.0	10533.0
1	5112	Block Group 1	25.075834	20.222447	40.6	37.1	0.0	0.0	18236.0
5	5105.02	Block Group 5	24.978760	37.638063	57.5	37.9	0.0	1.4	30963.0
2	5105.02	Block Group 2	23.349894	20.227111	22.7	59.4	0.0	0.0	18001.0
1	5305	Block Group 1	20.845921	34.592145	99.8	0.2	0.0	0.0	30827.0
3	5453	Block Group 3	19.708029	32.992701	97.7	2.3	0.0	0.0	38226.0
3	5109	Block Group 3	19.092332	12.989045	37.1	48.0	0.8	0.0	11006.0
3	5116	Block Group 3	17.300233	19.705198	26.5	70.2	0.9	0.0	21593.0
1	5109	Block Group 1	16.883117	26.298701	43.8	42.9	0.0	0.0	21776.0
3	5442	Block Group 3	16.167665	29.241517	97.9	0.0	0.0	0.0	26116.0

Suffolk County:

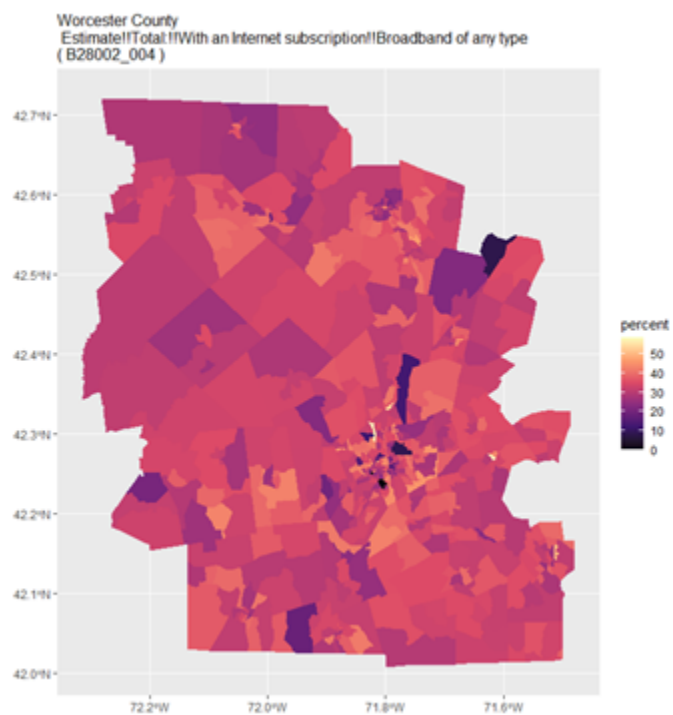
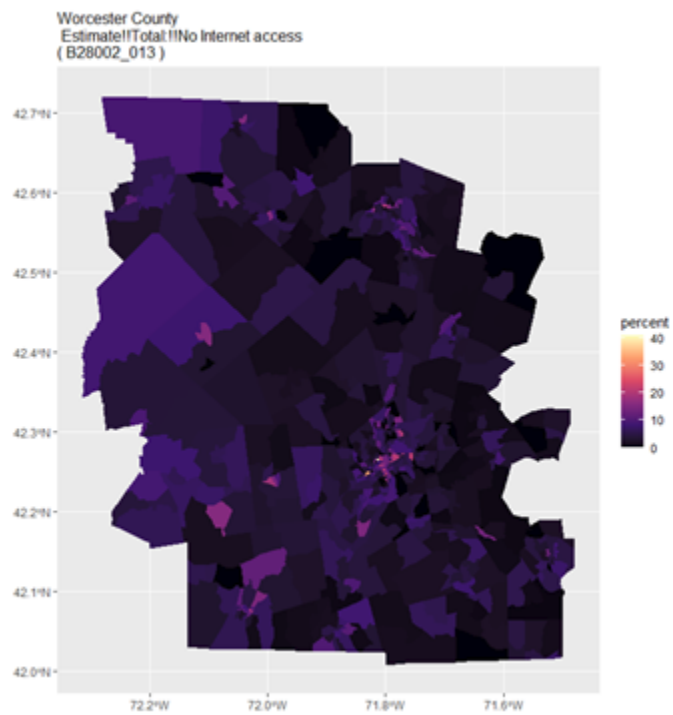




Suffolk County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
3	806.01	Block Group 3	32.846715	23.114355	18.0	64.5	0.0	1.5	15363.0
1	1404	Block Group 1	27.230047	18.153365	17.1	75.1	0.0	4.1	22420.0
1	813	Block Group 1	27.223354	19.008826	28.1	29.7	1.7	0.0	16593.0
2	705	Block Group 2	24.448217	38.539898	10.2	10.5	55.0	5.1	26183.0
4	922	Block Group 4	22.737307	24.944812	40.6	37.5	1.5	0.0	33179.0
2	817	Block Group 2	22.244290	27.706058	5.4	78.3	1.4	0.0	35839.0
1	901	Block Group 1	21.917808	17.727639	7.3	58.7	0.0	0.0	17142.0
3	916	Block Group 3	21.064815	23.379630	35.4	4.2	60.4	0.0	22623.0
1	817	Block Group 1	20.449438	13.483146	2.9	60.7	0.0	0.0	11740.0
7	1701	Block Group 7	19.322344	21.428571	68.5	4.7	2.7	0.0	23943.0

Worcester County:



Worcester County Demographics and Internet Information

	census tract	block group	% no internet	% internet with broadband subscription	% white	% black	% asian	% native american	mean income
2	7315	Block Group 2	40.677966	40.042373	58.3	27.3	0.0	0.0	18998.0
1	7312.03	Block Group 1	32.793807	24.349050	73.1	8.7	8.8	2.5	16195.0
1	7314	Block Group 1	21.336761	23.521851	77.9	12.5	0.0	2.2	11004.0
5	7326	Block Group 5	21.333333	20.266667	55.2	33.1	0.0	0.0	42966.0
1	7322.01	Block Group 1	20.886814	35.355893	83.0	0.0	11.1	1.3	41307.0
3	7316	Block Group 3	20.723918	33.002129	66.4	16.5	2.3	0.0	28977.0
6	7318	Block Group 6	20.070012	30.046674	64.9	14.5	7.1	0.9	13611.0
4	7304.01	Block Group 4	19.712352	20.558376	49.2	41.4	1.4	0.0	17595.0
2	7325	Block Group 2	19.413919	22.588523	77.7	6.1	4.8	0.0	17573.0
2	7107	Block Group 2	18.650794	28.174603	65.7	11.1	0.5	1.1	19979.0