Project Deliverable 3 (v1 Final Report)

All data should have been collected. All project questions should have been reviewed, answered, and submitted in a written document outlining findings as a PR. You will also be asked to submit the associated data and a README explaining what each label/feature in your dataset represents. Your team should meet with the client before this deliverable.

Checklist:

1. All data is collected
2. Refine the preliminary analysis of the data performed in PD1&2
3. Answer another key question
4. Attempt to answer overarching project question
5. Create a draft of your final report
6. Refine project scope and list of limitations with data and potential risks of achieving project goal
7. Submit a PR with the above report and modifications to original proposal

---

Up to now, all datasets have been collected and preprocessed.

Originally, we have 170731 small parcels, which are the parcels found and defined in many common Boston zoning viewers. To simplify the computation as well as to avoid going beyond the free usage cap of Google distance matrix API, we have successfully combined these small parcels into 13121 big parcels based on their geographical location. For example, if two houses are on different small parcels but their distance is less than 50 meters, they are considered to be within the same bigger parcel. We choose 50 meters as the parameter because it is a relatively short distance that can be covered in seconds. For different purposes, the parameter can be changed.

An example of merged small parcels is shown below. These small parcels are in the same block, and combining them does not affect the statistical value of our research.

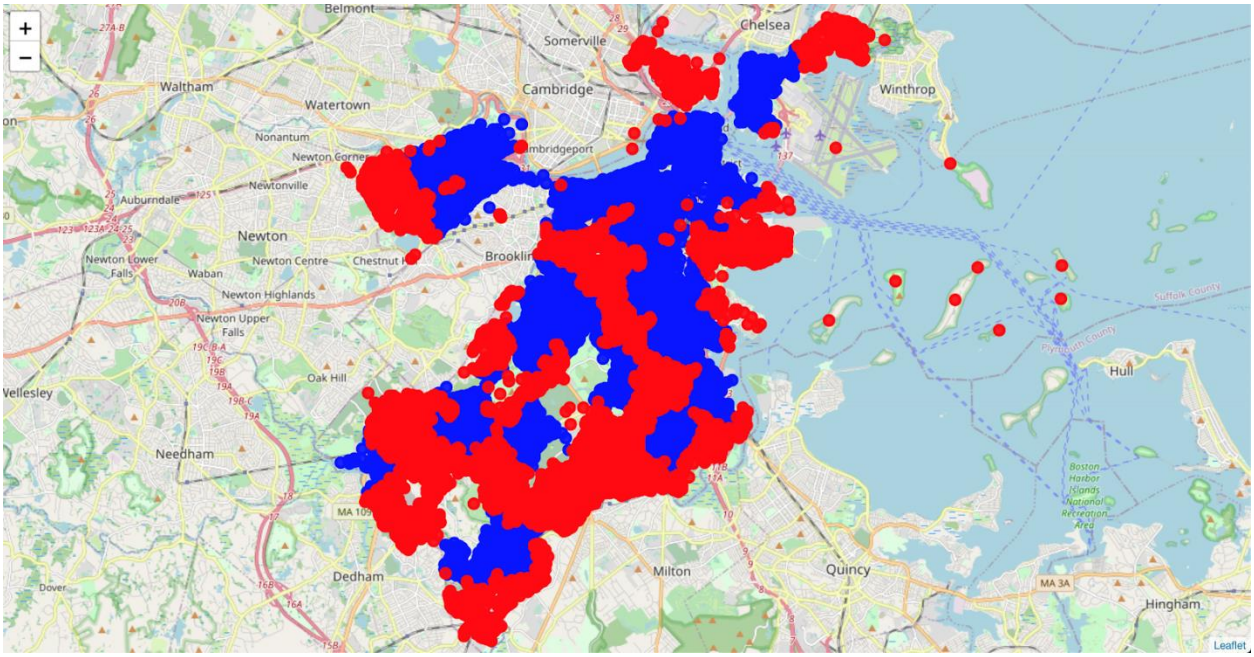
```
'27 -37 CHESTNUT ST #103, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #107, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #109, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #102, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #104, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #106, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #108, CHARLESTOWN,02129',
'27 -37 CHESTNUT ST #110, CHARLESTOWN,02129',
```

We have conducted research on 5 kinds of essential services: supermarket, grocery, hospital, health care and green space (including park). Statistically, we have found valid addresses for 68 supermarkets, 1876 groceries, 156 hospitals, 2493 health cares and 1027 green spaces in the City of Boston. Note that the number of hospitals may be too big because some hospitals have different buildings and these buildings are counted separately.

Using Euclidean distance, each big parcel is matched with several closest essential amenities of each kind. This redundancy compensates for the fact that the one with the smallest Euclidean distance may not be the one with the shortest road distance. Then we calculated the accurate road distances using Google map's Distance Matrix Api. Results are shown below.

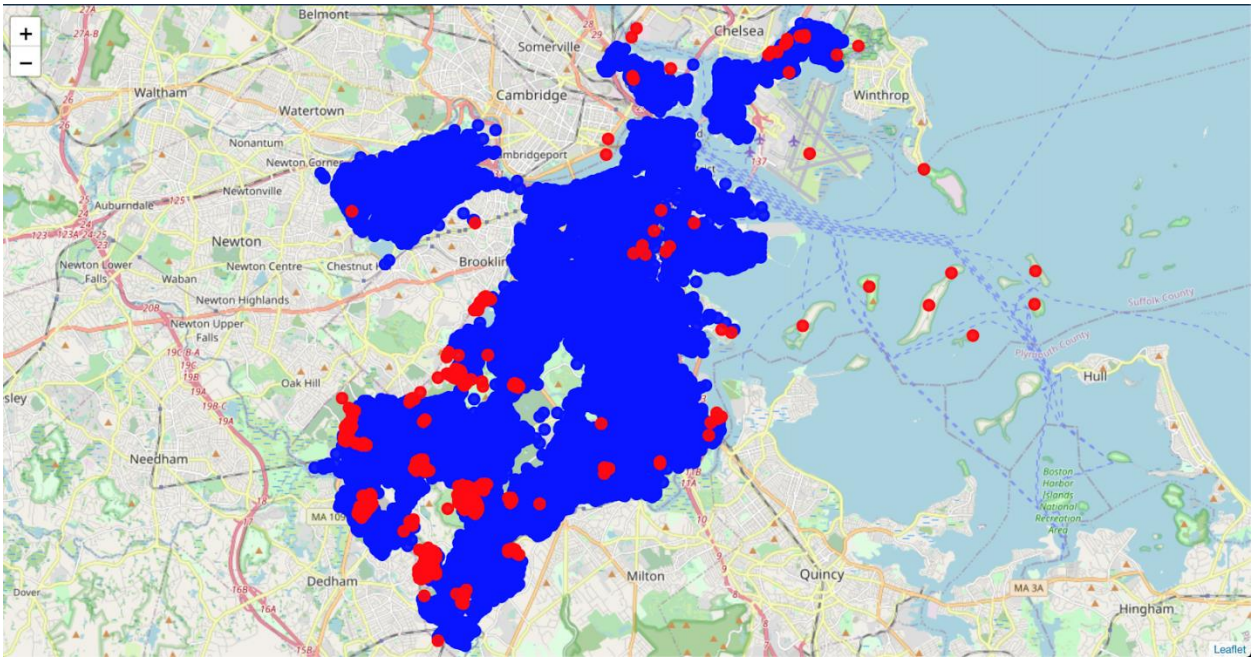
For each graph below, red denotes the areas that are not covered in 15-min walking while blue does the contrary.

I. Supermarket



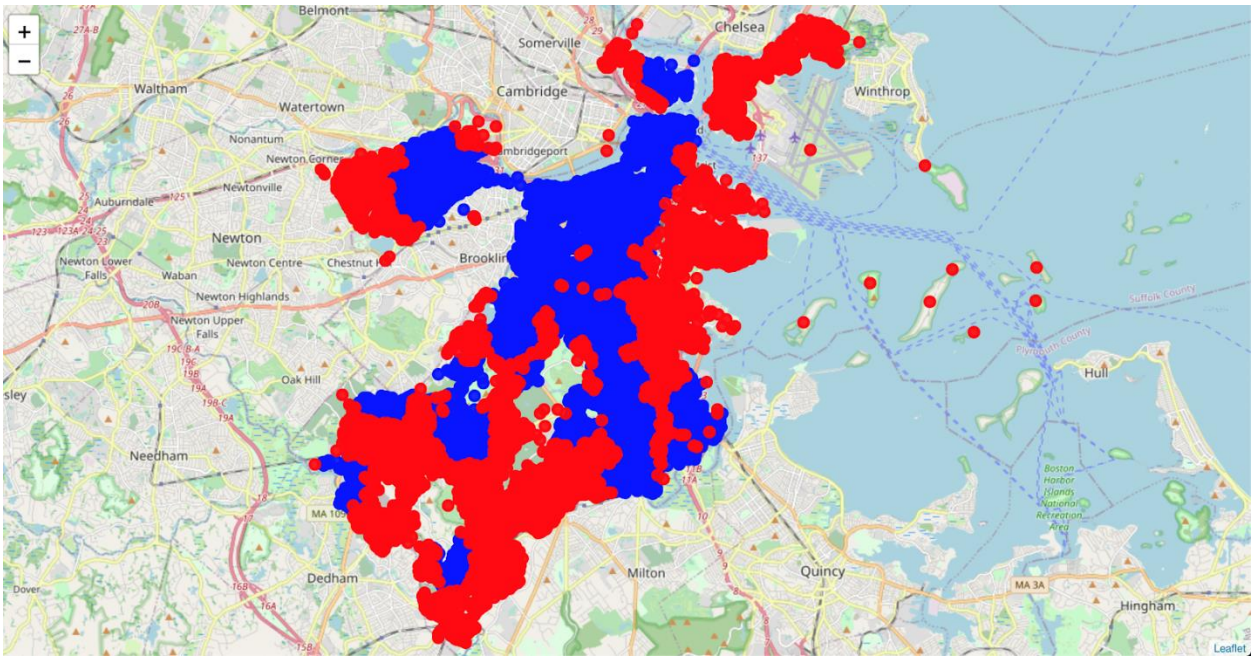
Percentage of parcels that are not covered is 0.468.

II. Grocery



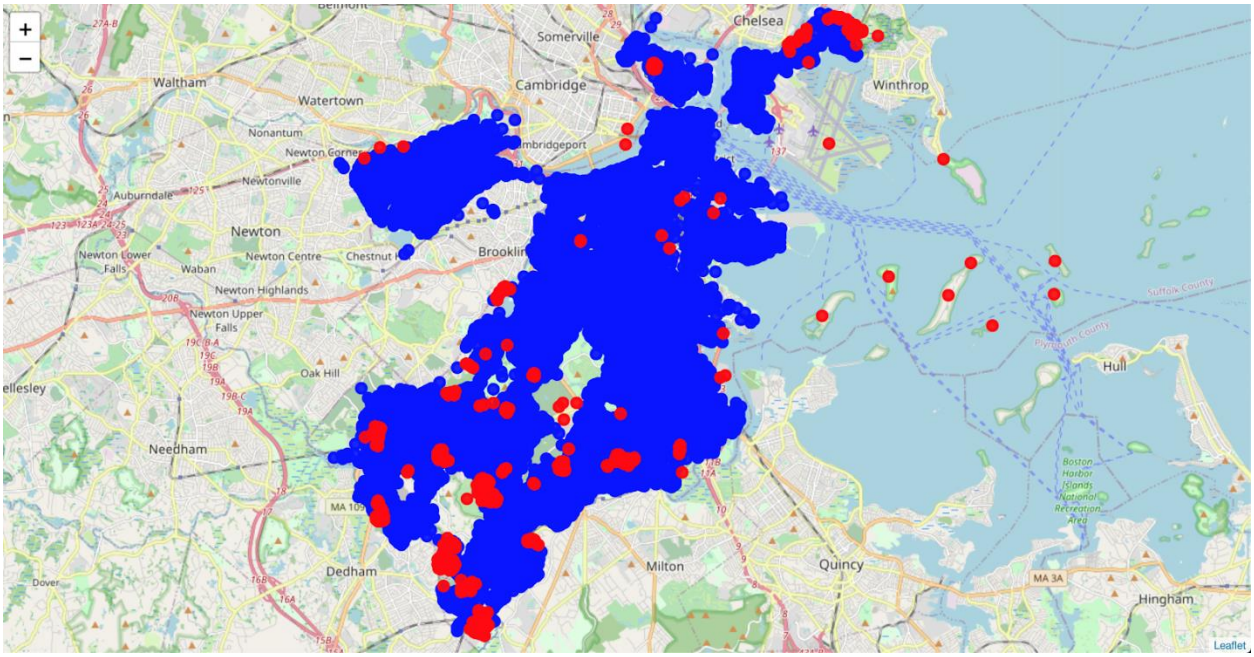
Percentage of parcels that are not covered is 0.0298.

III. Hospital



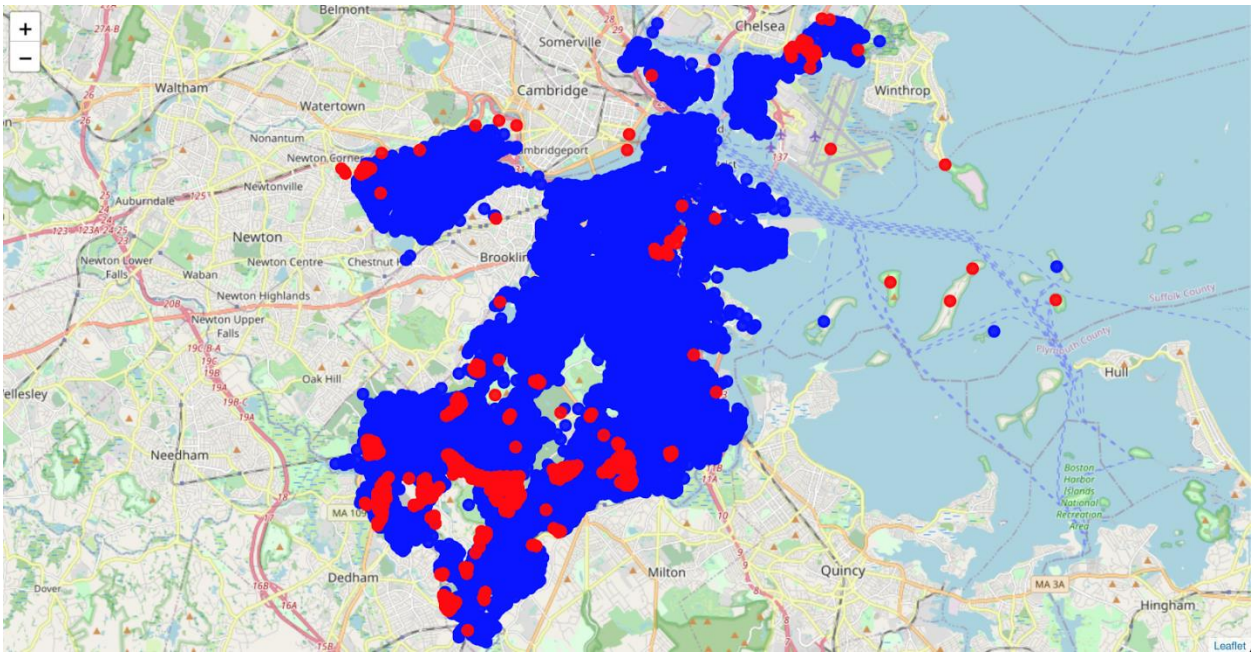
Percentage of parcels that are not covered is 0.472.

IV. Health Care



Percentage of parcels that are not covered is 0.0316.

V. Park and Green Space



Percentage of parcels that are not covered is 0.039.

Now, we are able to answer the 2 key questions in a better way.

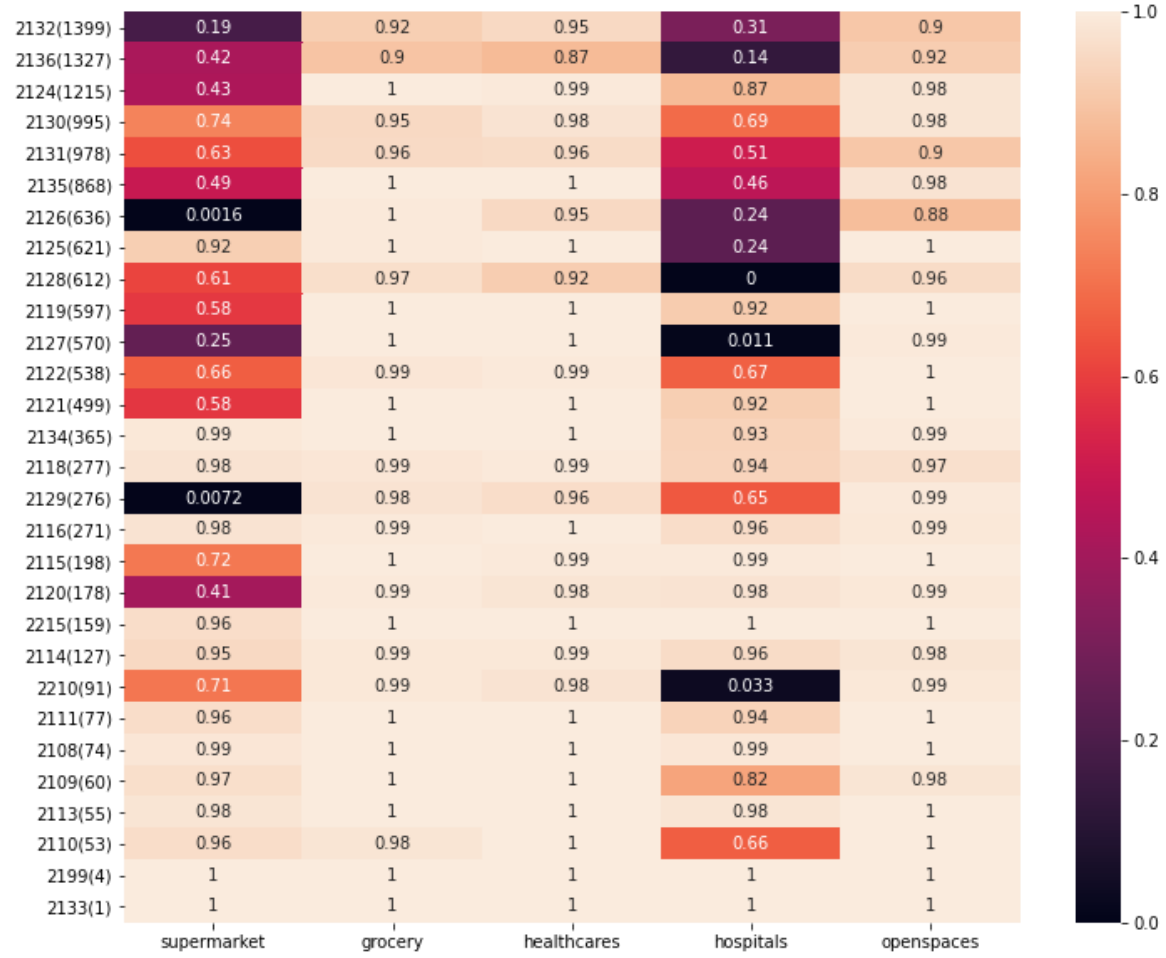
1. What percentage of residents are 15 minutes within essential amenities in a parcel of land?

As is mentioned above, the percentages of parcels that lack each kind of essential amenity (supermarkets, groceries, hospitals, health cares and green spaces) are 0.468, 0.0298, 0.472, 0.0316, 0.039.

2. Which areas of the city are underserved in terms of a lack of essential amenities?

Area analysis is done in terms of zip codes. For our dataset, we have 29 zip codes: [2136, 2111, 2131, 2116, 2129, 2132, 2127, 2126, 2122, 2113, 2215, 2118, 2124, 2125, 2114, 2119, 2121, 2130, 2110, 2128, 2199, 2115, 2120, 2134, 2135, 2210, 2109, 2108, 2133].

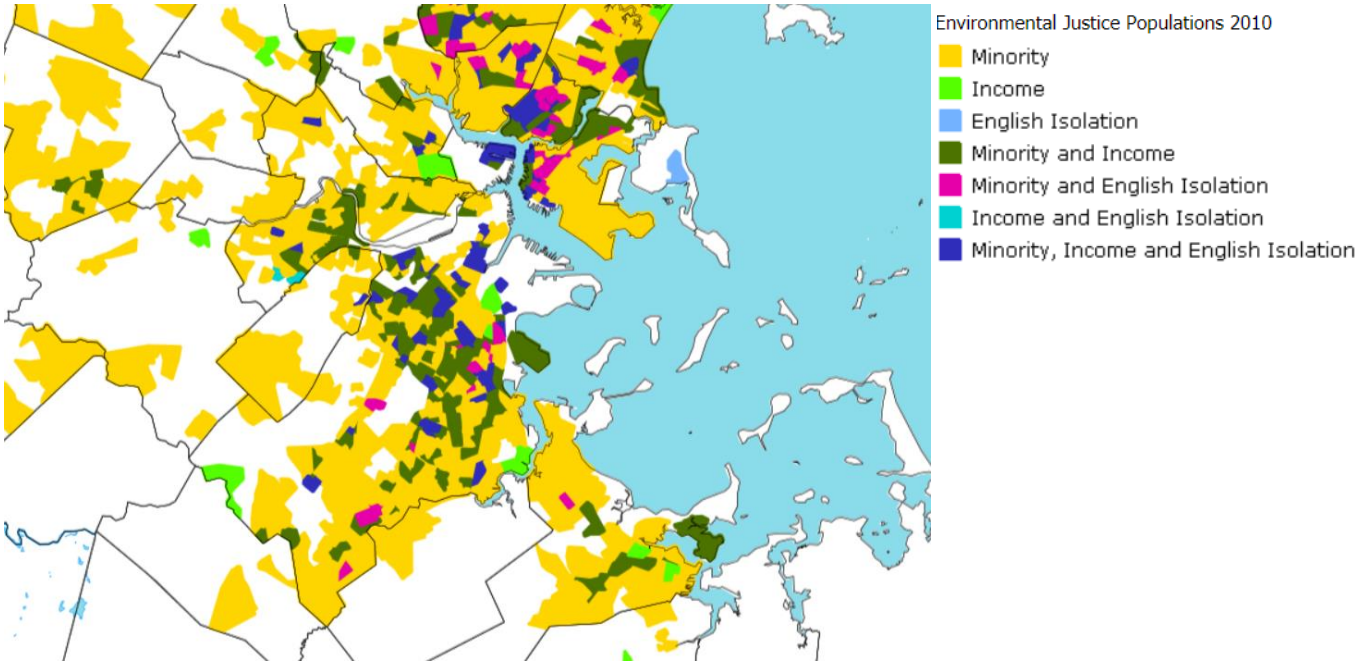
The heatmap of coverage rate of each service for each zip code is shown below, sorted in descending order of zip code size i.e. number of parcels included in each zip code. X axis represents services, Y axis is in the format of "*zip code(number of parcels included in this zip code)*".



It can be seen clearly that the most underserved services are supermarkets and hospitals. In this context, the most supermarket-underserved areas are 2132,2126 and 2129; the most hospital-underserved areas are 2136, 2128, 2127 and 2210.

Moreover …

Comparison to demographic map:



Given the race and income distribution of Boston city is so diverse, there does not seem to be a strong tendency with which people of minority or low income are underserved by certain kinds of essential amenity. In fact, in underserved parcels, the distribution of race and income shows diversity: both the majority and minority live there, with varied levels of income. So, whether people are underserved is not mainly based on their race or income, but which area they live.

Limitations, potential risks and possible improvement

- This analysis is done using parcels as the basic unit, but we don't have information about population in each parcel. Incorporating population data with parcel data can better reflect the real situation.

Appendix I: Service coverage rates of each zip code

	zipcode	number of parcels	supermarket	grocery	healthcares	hospitals	openspaces
0	2136	1327	0.419744	0.896006	0.873399	0.136398	0.920874
1	2111	77	0.961039	1.000000	1.000000	0.935065	1.000000
2	2131	978	0.630879	0.955010	0.960123	0.505112	0.904908
3	2116	271	0.977860	0.992620	0.996310	0.963100	0.988930
4	2129	276	0.007246	0.978261	0.963768	0.652174	0.992754
5	2132	1399	0.187277	0.924232	0.949964	0.309507	0.904217
6	2127	570	0.254386	0.996491	0.998246	0.010526	0.994737
7	2126	636	0.001572	0.995283	0.951258	0.242138	0.878931
8	2122	538	0.661710	0.986989	0.994424	0.665428	0.998141
9	2113	55	0.981818	1.000000	1.000000	0.981818	1.000000
10	2215	159	0.962264	1.000000	1.000000	1.000000	1.000000
11	2118	277	0.978339	0.985560	0.992780	0.938628	0.971119
12	2124	1215	0.428807	0.996708	0.992593	0.872428	0.983539
13	2125	621	0.924316	0.995169	0.998390	0.241546	0.998390
14	2114	127	0.952756	0.992126	0.992126	0.960630	0.984252
15	2119	597	0.582915	1.000000	1.000000	0.917923	0.998325
16	2121	499	0.581162	1.000000	1.000000	0.923848	1.000000
17	2130	995	0.739698	0.951759	0.977889	0.688442	0.982915
18	2110	53	0.962264	0.981132	1.000000	0.660377	1.000000
19	2128	612	0.609477	0.965686	0.919935	0.000000	0.957516
20	2199	4	1.000000	1.000000	1.000000	1.000000	1.000000
21	2115	198	0.717172	1.000000	0.994949	0.994949	1.000000
22	2120	178	0.410112	0.994382	0.983146	0.977528	0.988764
23	2134	365	0.989041	1.000000	1.000000	0.934247	0.994521
24	2135	868	0.489631	0.998848	0.997696	0.461982	0.980415
25	2210	91	0.714286	0.989011	0.978022	0.032967	0.989011
26	2109	60	0.966667	1.000000	1.000000	0.816667	0.983333
27	2108	74	0.986486	1.000000	1.000000	0.986486	1.000000
28	2133	1	1.000000	1.000000	1.000000	1.000000	1.000000