OVERVIEW

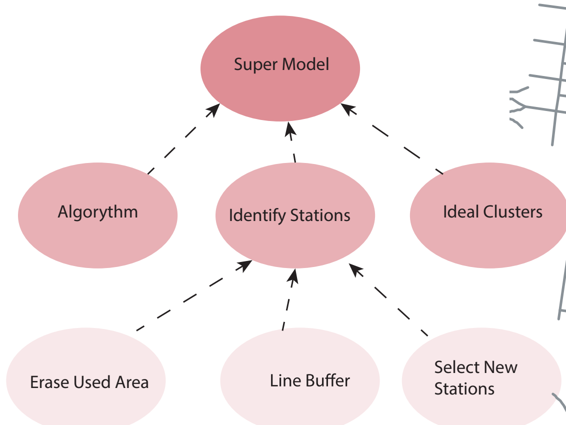
This report outlines the logic and methods used to build an instrument to model potential tax revenue to be earned by building a line of transit through a set of points representing households, each having a household attribute. In practice, one would need to put down a line, identify stations and then use an algorithm to identify the tax value expected from the housing points.

This model was built using a parcel dataset that was converted to points, each having something roughly equivalent to housing value in their attribute table. While this model was trained using a small subsection of a city, it can be applied to any geography. The subject area used is located between Brewerytown and the Strawberry Mansion area of philadelphia where housing values are low
and transit access is limited

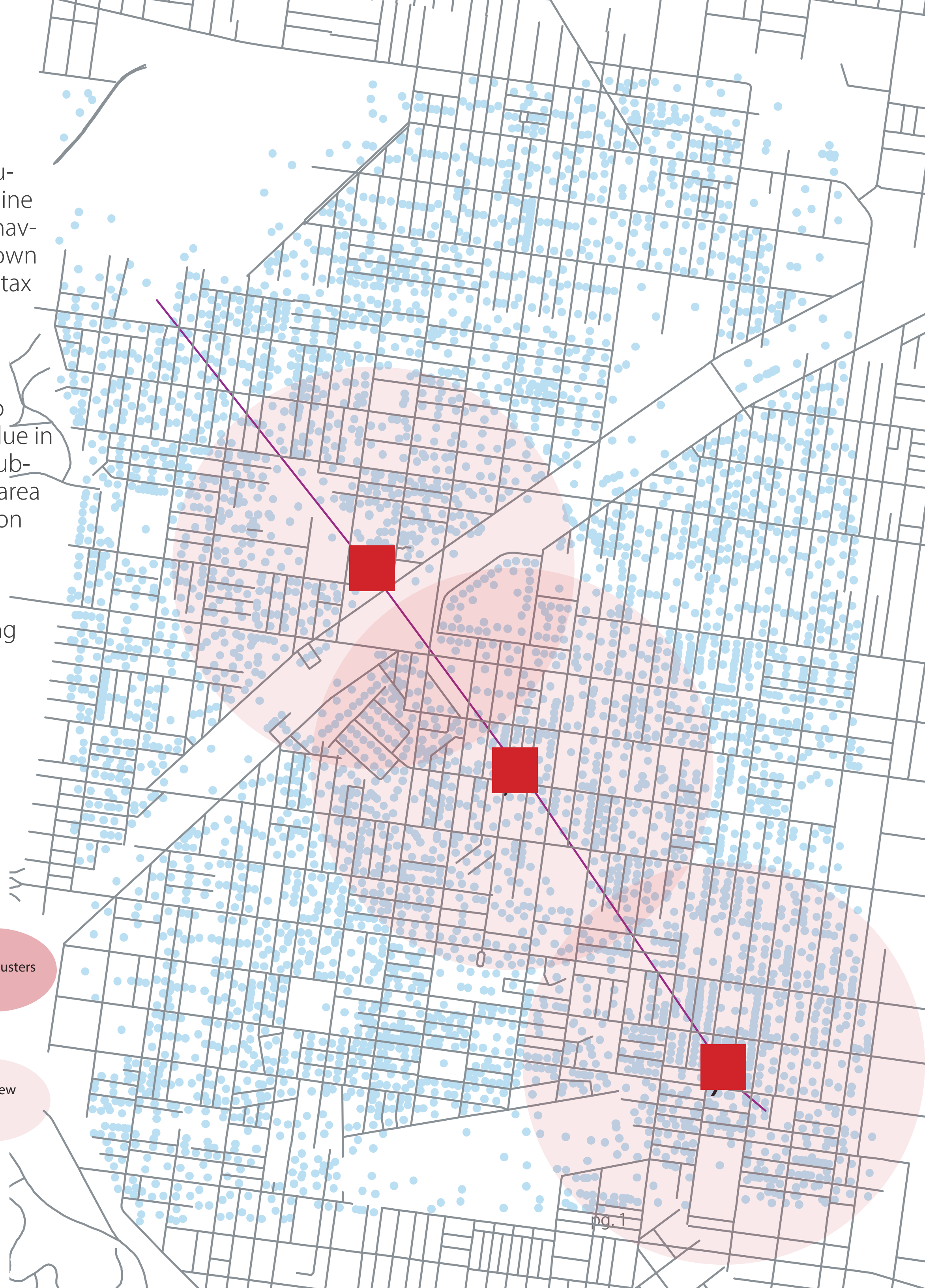
This description will walk through the peices of this model starting with the super model and working downwards.

The model has three key stages:

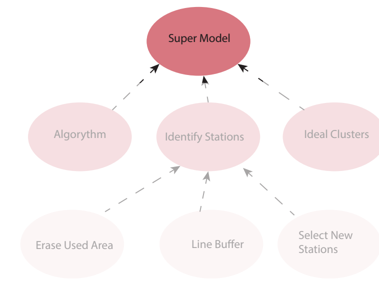
1. Identify areas of clustered activity
2. Identify the expected stations for the line.
3. Use a number of spatial variables to determine the tax value to be created, as a function of housing value increase.



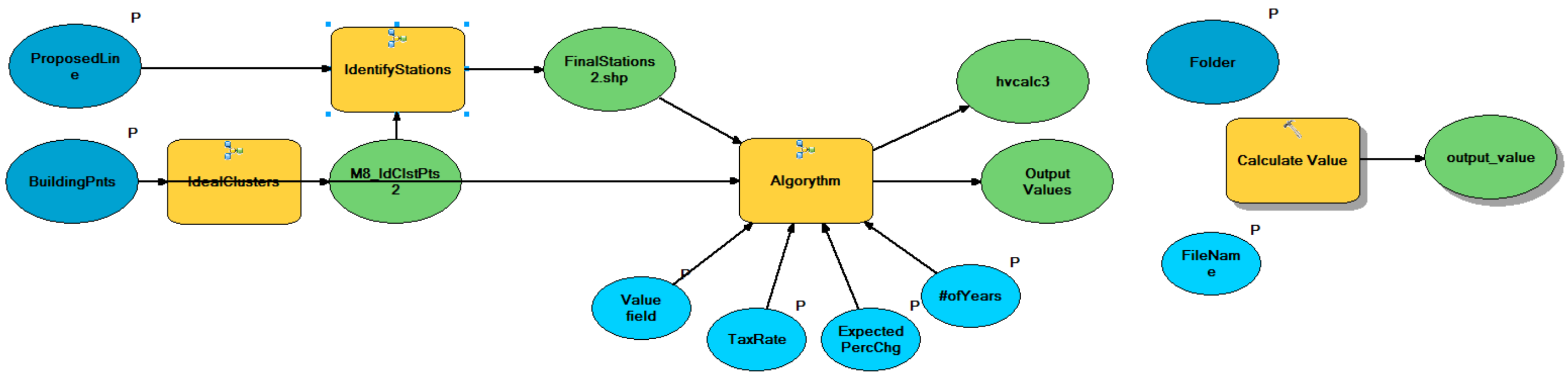
Station Identification process is broke up into three separate functions



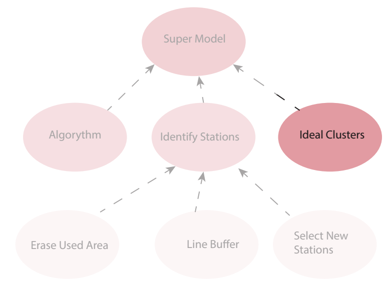
SUPER MODEL



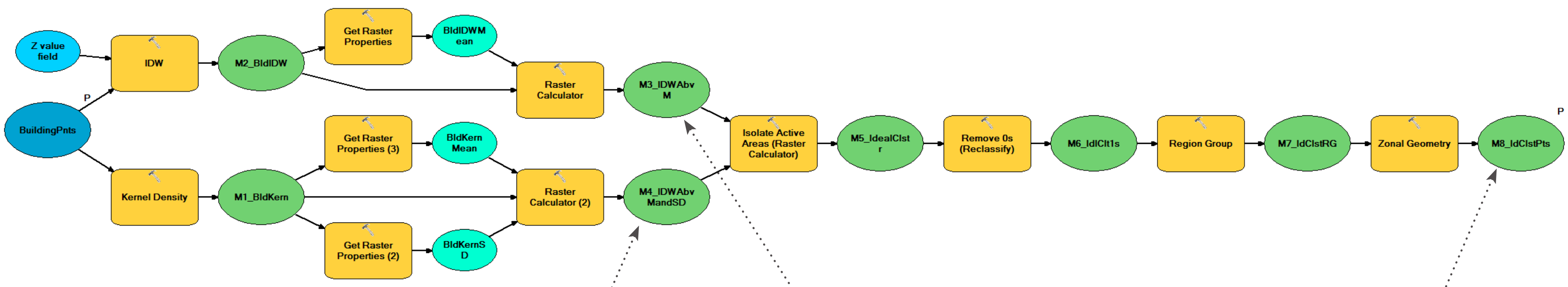
The super model brings in each of the various functions together. It takes the clusters and uses them and the line to identify appropriate stations, and then based on the station locations and building attributes as well as a number of key paramteters, the function calculates the expected tax revenue to be generated. The output values are a collection of expected tax revenues for each of years specified. For training this model, 10 years was used. This output was was then exported into a csv. The total expected generated revenue based on the proposed line and associated building footprint data is 21.5 million over a ten year period.



SUBMODEL 1: IDENTIFY ACTIVITY CLUSTERS



The model uses the points provided to create an IDW of housing value and a Kernel density of the points. Areas above the mean of the IDW output are isolated, as well as areas above one standard deviation above the mean of the kernel density output. The intersections of these rasters become the basis for what is identified as economic clusters.



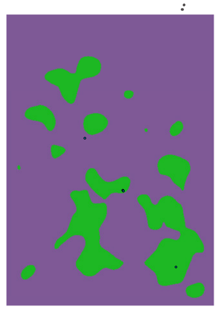


Figure 1: High IDW areas isolated

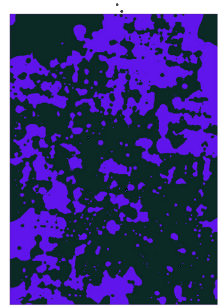


Figure 2: High Kernal Areas isolated

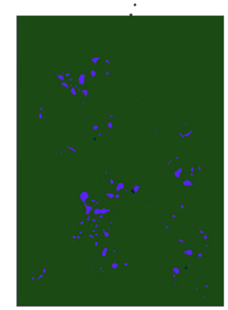
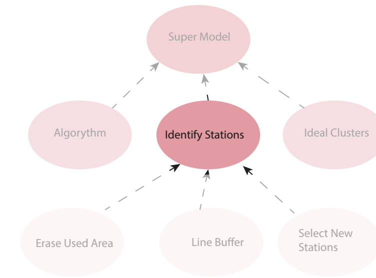


Figure 3: Final Identified Clusters

SUBMODEL 2: IDENTIFY STATIONS



To keep things organized, the process of identifying stations was broken into three sub-tasks. The model iterates until all stations are identified, which is determined by the ratio of station buffer to line buffer, which are outputs to the ***selecting new stations*** function. Once the ratio reaches 80%, the model stops iterating, meaning that 80% of the area around the line has been serviced with stations and no more stations are needed. The collected stations that are outputted by the model are merged into one point layer. The buffer of the point is used as a basis for removing the clusters used in order to identify a new station.

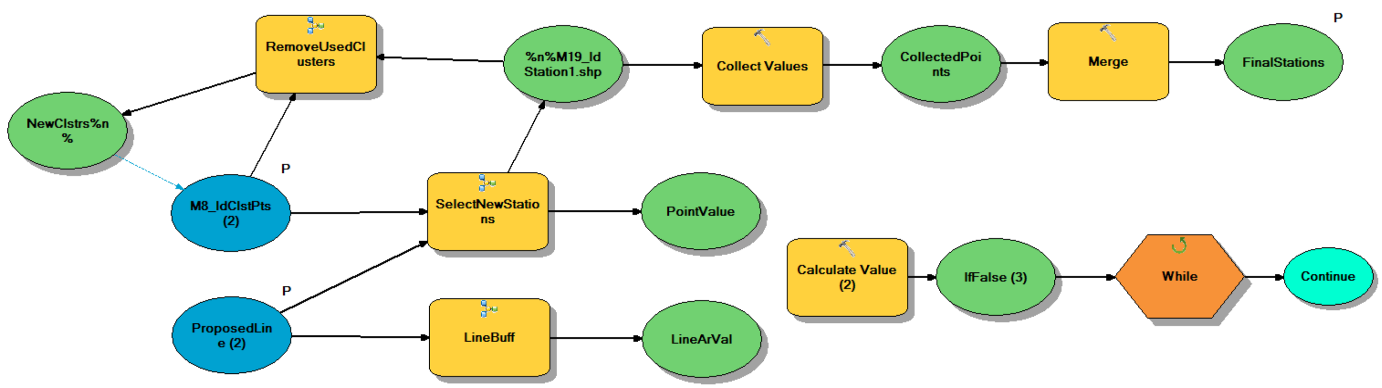
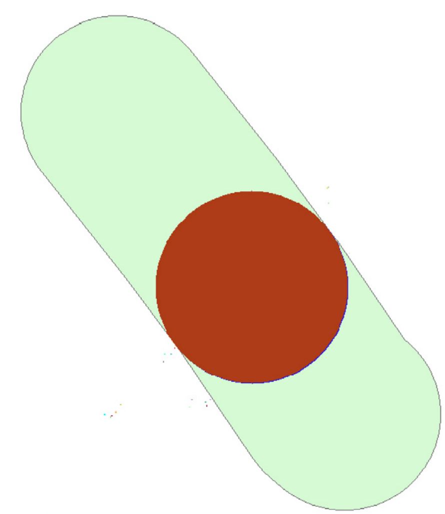
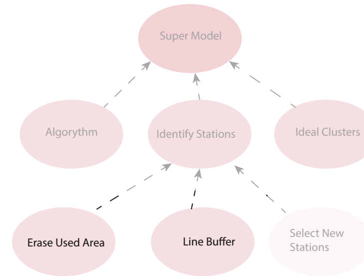


Figure 4: proportion covered over proportion to be

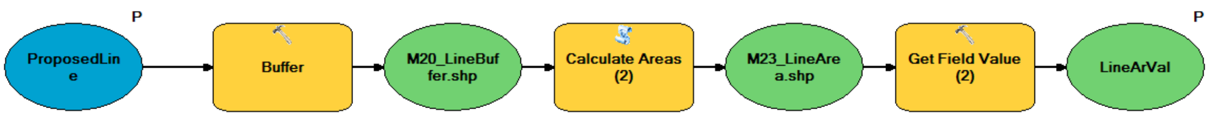


SUB-MODEL 2-1 Line Buffer and 2-2: Erase Clusters



In order to identify the point at which to stop iterating, a buffer around line segment is drawn and its area calculated.

The erase function uses the erase tool to remove the used clusters from the analysis. This output is then looped back into the select new stations function.



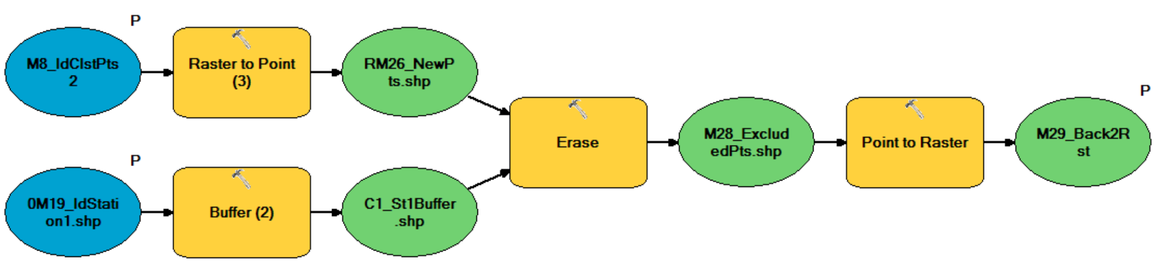
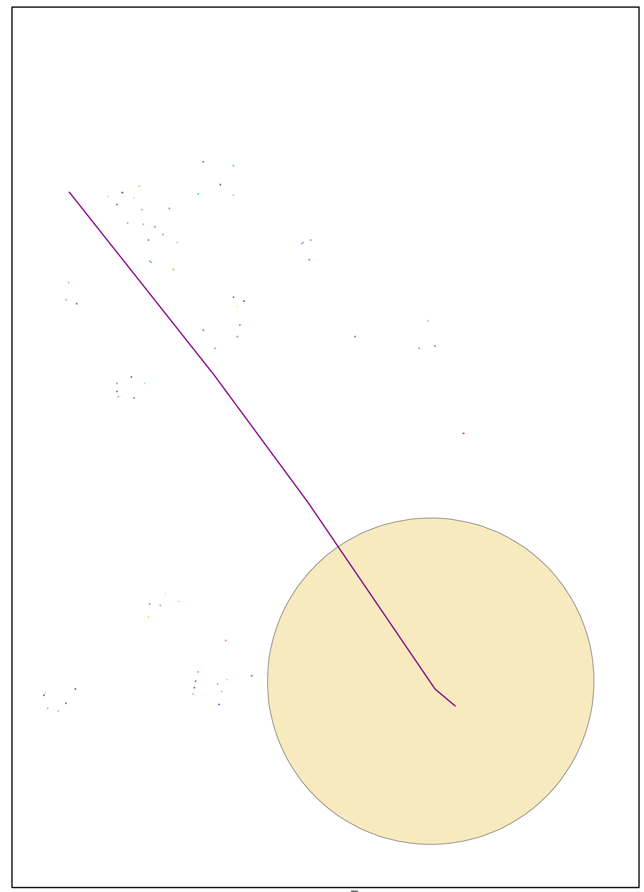
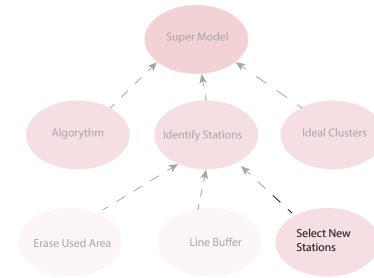


Figure 5: economic clusters erased by existing buffer



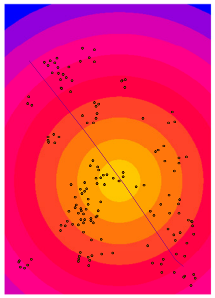
SUBMODEL 2: IDENTIFY STATIONS

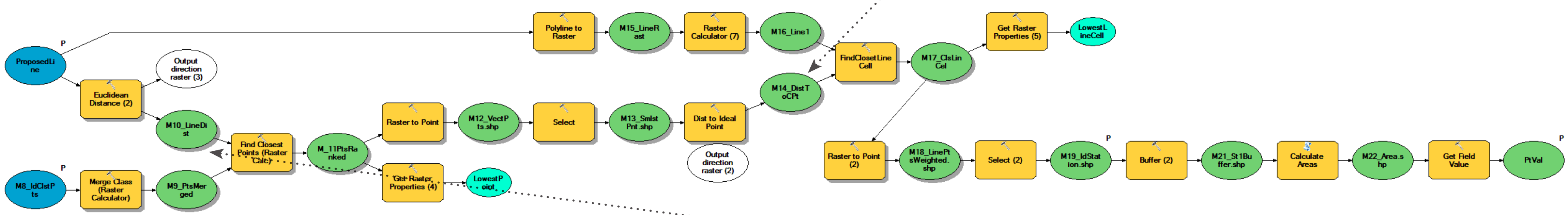


Once the intersections are identified the point closest to the line becomes the basis for the first station. A buffer is created around this station, and clustered economic areas associated with this line up to circle with a radius of 2,500 feet are removed. The next station is then created using a set of clusters with the previous clusters within the buffer omitted. The buffer used for this analysis is 2,500 which is the average distance between stations in Philadelphia, and likely a decent distance for stations generally. The stations are a little closer than expected, but like many factors in this model, the distance between stations can be tweeked.

The proximity to the line was determined using euclidean distance and raster calculator. The nearest part of the line was associated with the nearest cluster identify for each new station. The area with the lowest distance is selected.

Figure 6: Distance to station





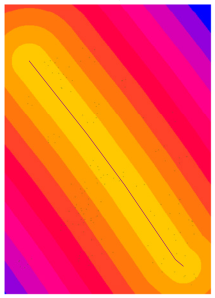
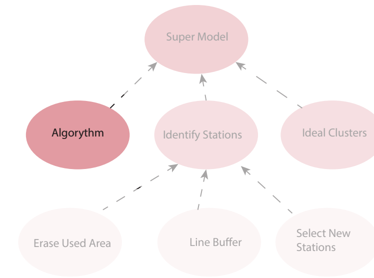
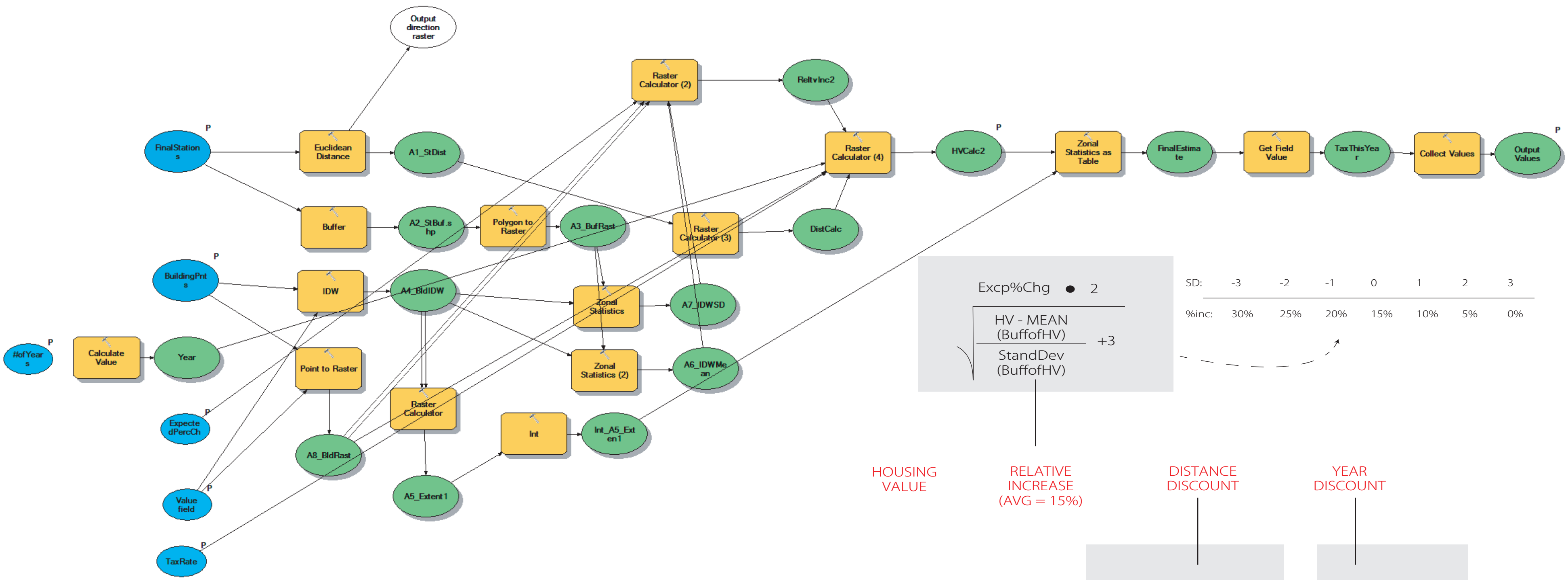


Figure 7: Distance to Line

SUBMODEL 3: VALUE CAPTURE ALGORITHM



There were 3 primary factors for house value appreciation associated with the installed stations. The first and primary factor is the overall expected increase in housing values for the areas adjacent. This increase is expected to be greatest for properties that are significantly below the average mean home value, while homes significantly above the mean home value are not expected to see as large an increase in housing value. This algorIthm was derived by reindexing the standard deviations and then devising an appropriate increment of percent increase by SD. The expected average increase is around 15%, which reflects the proportion the average person can save in their total annual



$$\frac{Excp\%Chg \bullet 2}{\sqrt{\frac{HV - MEAN(BuffofHV)}{StandDev(BuffofHV)} + 3}}$$

SD:	-3	-2	-1	0	1	2	3
%inc:	30%	25%	20%	15%	10%	5%	0%

HOUSING VALUE

RELATIVE INCREASE (AVG = 15%)

DISTANCE DISCOUNT

YEAR DISCOUNT

$$1 - (Distance * .005 * .03)$$

$$\frac{1}{Year\#} \bullet Tax\ Rate$$

Distance	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300
Increment	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5
	0.02	0.03	0.05	0.06	0.08	0.09	0.11	0.12	0.14	0.15	0.17	0.18	0.20
Discounted	0.99	0.97	0.96	0.94	0.93	0.91	0.90	0.88	0.87	0.85	0.84	0.82	0.81

Distance	1400	1500	1600	1700	1800	1900	2000	2100	2200	2300	2400	2500
Increment	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5
	0.21	0.23	0.24	0.26	0.27	0.29	0.30	0.32	0.33	0.35	0.36	0.38
Discounted	0.79	0.78	0.76	0.75	0.73	0.72	0.70	0.69	0.67	0.66	0.64	0.63

SUBMODEL 3: VALUE CAPTURE ALGORITHM, continued

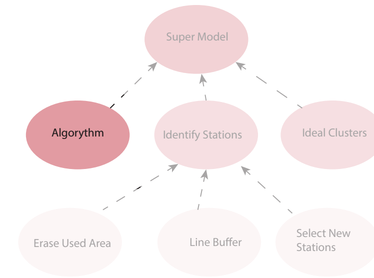
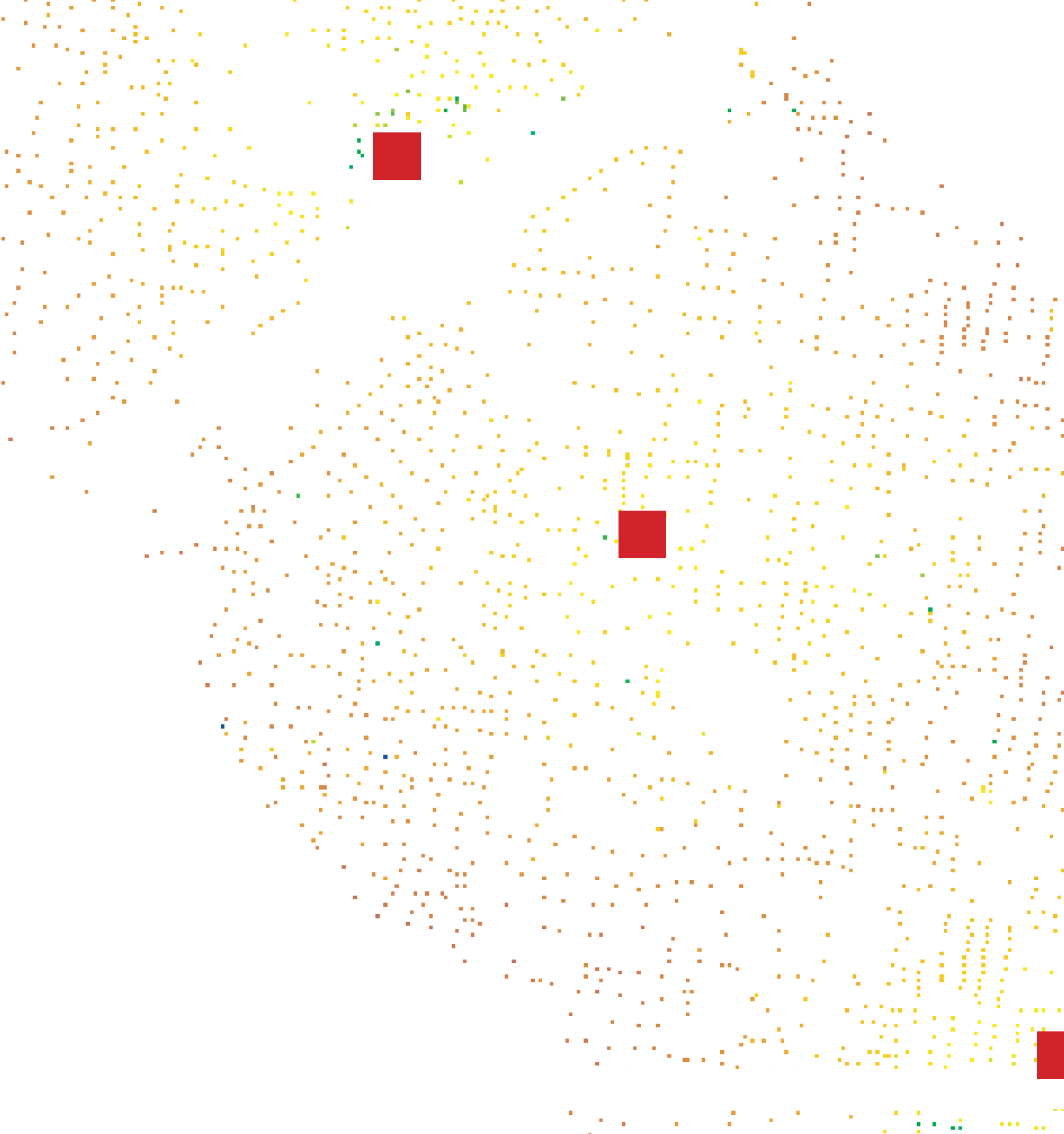


Figure 9: housing value expected increase



income on travel with transit. This expected increase is discounted based on proximity to the station, with the max discount reducing the expected increase by 47%, at 2,500 feet. Beyond 2,500 feet, it is assumed that there is little increase in housing values since that is beyond a reasonable walking distance. This model is iterated over a period of years specified in the model. The model was trained using ten years. This creates an exponential decrease in the overall housing value growth expected by the introduction of the transit line, as demonstrated in the below figure.

The total expected increase in tax revenue over the ten year period suggested by this model for this small area is 21.5 million dollars.

The outputs are a function of the parameters specified. This model provides a foundation that can be built for creating more useful land value modeling instruments.

Figure 8: tax revenue over time