Processing Documentation

We have a focal data set consisting of information about the demographics of libraries of Howard County. Out of our other two data sets ('Elementary_school_Districts' and 'Schools_Private'), I will be using the 'Schools_Private' dataset which contains the demographics of the private schools in Howard County. On combining these two datasets the new dataset will help us analyze how many libraries are located in the vicinity of schools.

I will follow a step wise descriptive approach as it will make it simpler to understand.

STEP 1: Choose the columns to process join operation

Since our chosen datasets have consistent data values, we will consider all columns except 'geom' in Schools_Private (because it does not give relevant data with respect to the question we are asking). So we will be considering the columns 'FID' (to help easily identify libraries from schools in the database), 'Name' (give basis of identifying an organization while mentioning it in the analysis report), 'Address' (location reference), 'City' and 'Zipcode' (the parameter which will help us analyze the location intersection between schools and libraries).

STEP 2: Create a new database with data from relevant columns from both datasets.

There are many methods to do this task. If we use R, the script required is the following:

```
>
>
>
> d=read.csv('C:/Users/sanjn/Downloads/Libraries.csv')
> a=read.csv('C:/Users/sanjn/Downloads/Schools_Private.csv')
> merge(d, a, by = "Zipcode", all = T)
   Zipcode                                FID.x              Name.x               Address.x       City.x                                  geom.x
1    20723 Libraries.fid-e93e948_1508c03aa40_231c              Savage           9525 DURNESS LN       Laurel POINT (1358731.9977323646 533992.2972172729)
2    20763                                 <NA>                <NA>                      <NA>         <NA>                                    <NA>
3    20794                                 <NA>                <NA>                      <NA>         <NA>                                    <NA>
4    21029                                 <NA>                <NA>                      <NA>         <NA>                                    <NA>
5    21042 Libraries.fid-e93e948_1508c03aa40_231b        Miller Branch         9421 FREDERICK RD Ellicott City POINT (1358095.7461026195 584929.8885943966)
6    21042 Libraries.fid-e93e948_1508c03aa40_231b        Miller Branch         9421 FREDERICK RD Ellicott City POINT (1358095.7461026195 584929.8885943966)
7    21042 Libraries.fid-e93e948_1508c03aa40_231b        Miller Branch         9421 FREDERICK RD Ellicott City POINT (1358095.7461026195 584929.8885943966)
8    21043 Libraries.fid-e93e948_1508c03aa40_2318 HC HISTORICAL SOCIETY            3725 PARK AVE Ellicott City POINT (1369461.6468812667 583291.8479450485)
9    21043 Libraries.fid-e93e948_1508c03aa40_2318 HC HISTORICAL SOCIETY            3725 PARK AVE Ellicott City POINT (1369461.6468812667 583291.8479450485)
10   21043 Libraries.fid-e93e948_1508c03aa40_2318 HC HISTORICAL SOCIETY            3725 PARK AVE Ellicott City POINT (1369461.6468812667 583291.8479450485)
11   21043 Libraries.fid-e93e948_1508c03aa40_2318 HC HISTORICAL SOCIETY            3725 PARK AVE Ellicott City POINT (1369461.6468812667 583291.8479450485)
12   21044 Libraries.fid-e93e948_1508c03aa40_231e Howard County Central 10375 LITTLE PATUXENT PKWY     Columbia  POINT (1352482.742419602 562493.6777206546)
```

I noticed that although the task is done, the redundancy of zip codes (which is the basis for our analysis) is not conspicuous.

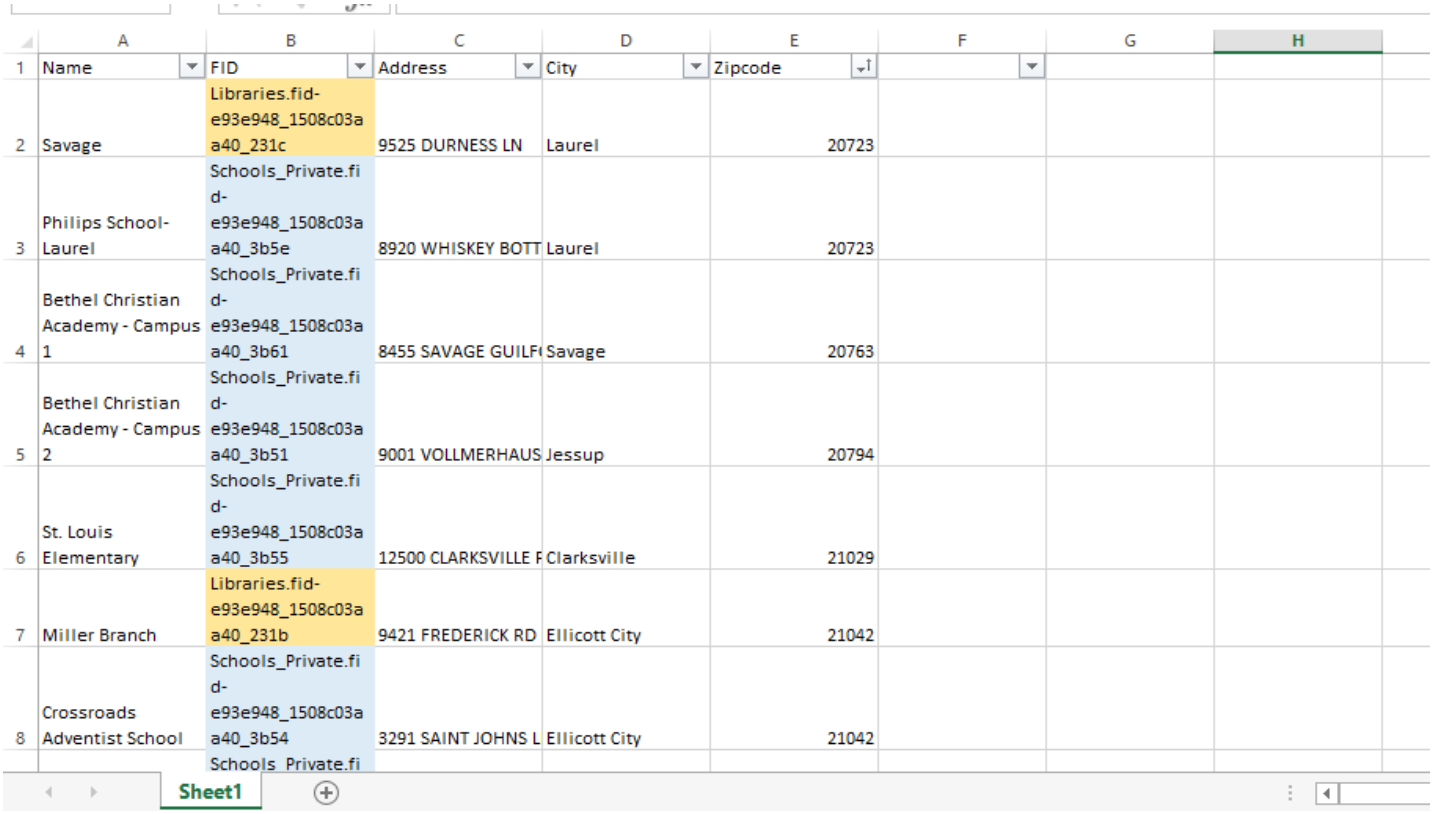
That's why I will be using Excel for creating the new Dataset. For this task I made a copy of the bigger dataset (Private_schools) and I concatenated data values from the 'Libraries' dataset to this copy. We can delete the 'geom' column as discussed earlier.

Step 3: Filter, and alter data to aid in superficial analysis.

We have our new dataset ready for analysis. Firstly I color coded the institutions based on type (i.e. in the FID column, I colored the schools to be blue and the libraries to be yellow) this will give us a better visual understanding as soon as we look at the data.

We apply a filter to the 'Zipcode' column by selecting the column and selecting the filter option in the toolbar above. I chose to sort them in ascending order, so now we can easily identify redundancy and by the color code easily identify how many schools have a library in the vicinity.

Here is a screenshot of the filtered dataset.



	A	B	C	D	E	F	G	H
1	Name	FID	Address	City	Zipcode			
2	Savage	Libraries.fid-e93e948_1508c03aa40_231c	9525 DURNESS LN	Laurel	20723			
3	Philips School-Laurel	Schools_Private.fid-e93e948_1508c03aa40_3b5e	8920 WHISKEY BOTT	Laurel	20723			
4	Bethel Christian Academy - Campus 1	Schools_Private.fid-e93e948_1508c03aa40_3b61	8455 SAVAGE GUILF	Savage	20763			
5	Bethel Christian Academy - Campus 2	Schools_Private.fid-e93e948_1508c03aa40_3b51	9001 VOLLMERHAUS	Jessup	20794			
6	St. Louis Elementary	Schools_Private.fid-e93e948_1508c03aa40_3b55	12500 CLARKSVILLE F	Clarksville	21029			
7	Miller Branch	Libraries.fid-e93e948_1508c03aa40_231b	9421 FREDERICK RD	Ellicott City	21042			
8	Crossroads Adventist School	Schools_Private.fid-e93e948_1508c03aa40_3b54	3291 SAINT JOHNS L	Ellicott City	21042			
		Schools_Private.fi						

Sheet1

For a superficial understanding this dataset gives us a good level of visualization about the library's near schools.