1) This data quality analysis will be assessing the accessibility of three schools in the 01602/01609 zip code of Worcester, Massachusetts. The accessibility of schools is important to consider as most students travel to school by foot or automobile through winding and confusing residential areas. The positional accuracy of the data should be within 100 feet, as less accurate data could lead students and drivers down the wrong roads, increasing their walking and driving time. A positional accuracy of within 100 feet would also be important for determining what residential neighborhoods should be included on school bus routes, based on total distance from the starting point – the bus stop – to the ending point – the school.

2) In the images below, the MassGIS road centerlines appear in orange, the Census TIGER lines in green, and the StreetMap USA lines in pink.

MassGIS road centerlines:
The orange lines represent the road centerline data from MassGIS, produced by the Executive Office of Transportation. Out of all three road centerline data sets, MassGIS data best follows the actual path of the roads. The MassGIS lines generally fall in the middle of the roads depicted by the aerial photograph. In many cases, the Census TIGER data and the StreetMap USA data do not detect smaller, local roads, as seen on the right side of the map below, but MassGIS seems to have data for nearly every road in this neighborhood.

StreetMap USA road centerlines:
In general, the StreetMap USA road centerlines, depicted on the map below in pink, are more extensive than the Census TIGER centerline data but less positionally accurate. In addition, they are not nearly as accurate as the MassGIS centerlines and they are far less likely to fall in the center of the road, often appearing on the edge of the road or not on the road at all, and inconsistent in their placement. In the map below, one can see that MassGIS and StreetMap USA both depict a T-shaped intersection. (The Census TIGER data does not depict the intersection at all.) While the orange MassGIS lines are in the center of these small
local roads, the pink StreetMap USA lines appear to be on top of residences 212.55 feet away from the actual location of the road.

Census TIGER road centerlines:
In comparison to the MassGIS road centerlines, the Census TIGER lines (in green) are less positionally accurate. In the map above, the Census TIGER lines are between 52 to 95 feet away from the more accurate MassGIS roads. However, in this neighborhood, the Census TIGER centerline data are more accurate than the StreetMap USA data, which shows the roads to be even farther away than they really are (over 200 feet off). However, both the MassGIS data and the StreetMap USA data are more extensive than the Census TIGER data. The Census TIGER data do not show all of the local roads, such as the T-shaped intersection in the map above (the intersection is represented by both MassGIS data and StreetMap USA data in the map above).

From this analysis, it is clear that the MassGIS road centerline dataset is the most useful for this project, as the data is most positionally accurate and complete.

3) The Census TIGER water dataset was very inaccurate. In fact, it did not recognize one body of water in the neighborhood, although there are several. Therefore, I included the DEP wetlands data layer, provided by MassGIS, for additional comparison. The MassGIS hydrography data layer is depicted in the maps below with a blue line, whereas the DEP wetlands data layer is shown as blue polygons.

The MassGIS dataset is more accurate and extensive than the DEP wetlands dataset. As seen in the map below, the MassGIS outline of the pond (blue line) is more similar to the pond shown in the aerial photograph (in dark green). The DEP wetlands data (blue polygon) is less accurate, and the pond is shown to be much smaller than it actually is. For example, on one side of the pond, the DEP wetlands data boundary and the MassGIS data boundary differ by over 93 feet.
However, none of the datasets account for all of the hydrography in the area. Because I live in this neighborhood of Worcester, I know that there is a stream in the area depicted below. The stream is located approximately where I drew the blue curvy line. It is to the right of a school and at the end of a dead-end road. None of these datasets account for such small bodies of water, even though the metadata suggests that they do.

4) In order to assess the positional accuracy of the MassGIS datasets, I used the US National Map Accuracy Standard and read through the online metadata at Mass.gov, as it was the most detailed metadata available.

• The MassGIS road centerline data has a scale of 1:5,000, meaning that the map accuracy should be about +/- 13.33 feet. This seems plausible, as the MassGIS road centerlines were the most accurate of the three road datasets.
• The positional accuracy of the Census TIGER road centerline data could not be quantitatively assessed as the online metadata says that scales of either 1:100,000 or 1:24,000 were used to make the data, meaning that the accuracy of the data varies.

• The accuracy of the StreetMap USA road centerline data is no more accurate than 1:100,000, meaning that the road centerline accuracy should be about +/- 166.67 feet. In the map below, the road centerlines are slightly less accurate than predicted, with the StreetMap USA roads being about 212.5 feet away from the road centerlines depicted by the aerial photograph.

• The MassGIS hydrography data layer uses a scale of 1:25,000; therefore, the map accuracy should be about +/- 40 feet.

• The DEP wetlands data is at a scale of 1:12,000; therefore, the map accuracy should be +/- 33 feet. However, the first map shown in Question 3 above demonstrates that it is unlikely that this data is +/- 33 feet accurate, given that the water body depicted in the aerial photograph covers a much larger territory.

• The accuracy of the Census TIGER hydrography data could not be quantitatively assessed as several scales (including 1:100,000 and 1:24,000) are provided, making it difficult to quantify the accuracy.

• The Land Use MassGIS dataset is at a scale of 1:25,000; therefore, the map accuracy should be around +/- 40 feet.

• The MassGIS school data uses both a scale of 1:5,000 and 1:25,000, based on available data. Therefore, the map accuracy should range from about +/- 13.33 feet to +/- 33.33 feet, depending on what scale was used to create the schools in this study area.

• While the scale used in the MassGIS library dataset is unknown, the library metadata explains that the libraries were geocoded based on address. They also used online mapping resources and made phone calls to the libraries in order to achieve a 100% match rate. Additionally, the one library included in this particular study area is within a school; therefore, it is likely that it meets the accuracy range of the MassGIS school data (about +/- 13.33 feet to +/- 33.33 feet).

• The prime forest land dataset from MassGIS was based on digital soil data, and was not extracted from paper maps. MassGIS does not provide a scale for the prime
forest dataset. It can be assumed that this data is more accurate than data that was digitized from a paper map, as no additional errors were introduced from that process.

5) The four optional data layers that I chose to best assess the accessibility of schools in this area of Worcester were: 1) schools, 2) libraries, 3) prime forest land, and 4) land use.

- The school data appears to be positionally accurate in relation to the schools shown in the aerial photograph. However, the schools are depicted as points, not polygons. Therefore, they are much smaller than the actual school, as you can see in the map below. The school data from MassGIS, represented by the purple dot, does not allow for an easy assessment of the best route to use to get to this school. Because the school shown in the map below is in my neighborhood, I know that the main entrance is in the area enclosed by the red circle, and there is a dead end street at the red X in the top left corner. It would be better if this data was in polygon format and showed the main entrance of the school, as that would better represent school location in relation to local streets.

![Aerial photograph of school location](image)

- The library data within this area of Worcester seems to be fairly positionally accurate. However, the library included in this area of Worcester is within a school and is also a point, not a polygon. Therefore, it does not show the exact location of the library and it is most likely much smaller than the actual library itself. It also does not tell you that the library is not open to the general public, as it is on the property of a private, all-girls Catholic high school. In addition, I know that another elementary school within the area, Flagg Street School, has a library. The Flagg Street School library is small and only for the children and faculty of the elementary school. However, I think it is neglectful to not include it in this data set.

- Overall, the prime forest land MassGIS data set is moderately accurate. It accounts for all of the forested areas in the neighborhood, covers the majority of the forested territories, and depicts forests in the correct locations in relation to local streets. However, the prime forest land data set also includes a few houses, parking lots, and other non-forest areas. One house (in the red circle in the upper left hand side) and one parking lot (in the red circle in the bottom right) can be seen in the aerial photograph in the map below. This is simply inaccurate forest data, as such territories should not be included in forest area datasets. In addition, on the left side of the map one can see forest land that is not included in this MassGIS data set. As a resident of
this neighborhood, I know that this forested area is indeed forest, and should probably be included in the data set.

- The land use MassGIS dataset is quite accurate. It accurately shows high density residential areas, medium density residential areas, forests, one body of water, and three “urban public” areas, or schools. The boundaries of each of these areas seem to be slightly imprecise, as some boundaries are oddly rectangular, while others seem a little vague. However, my knowledge of the area allows me to believe that this imprecision is minor, as the high density residential areas are well-defined and the water body and forest areas match those of the aerial photograph.

6) These four optional data layers would be more appropriate for the assessment of accessibility to local schools if they were more positionally accurate.

- The point data for schools, although positionally accurate, is so much smaller than the actual school that it is difficult to determine from which side is best to access the school. This could leave students and parents spending excess time trying to enter the school building at its appropriate entrance.

- The library data is incomplete, as it does not show the library at Flagg Street School. While the one library shown at Notre Dame Academy is positionally accurate, there is no data for the other library.

- The prime forest MassGIS data is limiting in that it covers a parking lot and a minor road (as seem in the second map in Question 5). If a driver were to look at this dataset to determine how to access the school, he would not know that a direct driving route as well as a parking area are covered by this positionally inaccurate forest data.

- The land use MassGIS dataset is the most appropriate data set for this assessment out of all four optional layers. It accurately shows the different land uses, which allows walkers or drivers to select their ideal route to school. A walker may choose to walk through a high density residential neighborhood because it is likely that there are sidewalks and less commercial traffic. From this data, a driver is able to see where there is forest and water and avoid those areas. The urban public areas are clearly defined, including their parking lots and green space. The map below is a good example of how this land use data is appropriate for this assessment. The MassGIS school data is represented by the purple dot in the center, yet the land use data is much more positionally accurate, showing the school to be much larger (the entire light blue area with small polka dots), extending all the way to the right edge of the map. The forested area in green surrounding the left and the bottom side of the school suggests that it is best to enter the school property from the right side. (In
fact, the MassGIS road data, shown with an orange line, is inaccurate as there is no road that connects from the left side of the forest to the right side of the forest. This road is actually a dead end that does not lead to the school.)

7) The completeness of each of the nine data sets varies.
   - The MassGIS road centerlines data set seems to be fairly complete, although eleven of the streets in the area have missing Street Name information. It covers the study area, and all relevant features (road centerlines shown in orange in the map below) are present.
   - The StreetMap USA road centerlines data set is less complete, in that the roads do not have information such as Street Name or Road Type. While the relevant feature (road centerlines) is present, the StreetMap USA road centerlines (shown in pink) do not cover the study areas as extensively as do the MassGIS road centerlines (shown in orange), as seen in the map below. This incomplete StreetMap USA road centerlines data could make for a confusing assessment of school accessibility.
   - Although it does not appear this way on the map, the Census TIGER road centerlines data is more complete in terms of attribute information than the StreetMap USA
data. The attribute information is not fully complete, but it provides street name and street type information for some of the streets (about 50%). The Census TIGER road centerline data does not cover the study area very well. The Census TIGER roads are represented in green, but one will notice that they do not even appear in the map above, as there is no Census TIGER road data for the local streets near the school.

- The Census TIGER hydrography dataset provides complete attribute information for all features, but it does not cover the study area at all. No relevant features are present.
- The MassGIS hydrography data covers the study area. It recognizes three bodies of water in the area, which is not all of the water bodies, but it is far more than what is depicted by the Census TIGER hydrography data. The attribute information is complete except for the Public Water Supply ID Number of Downstream Public Surface Drinking Water Supply (PWSID). There are no PWSIDs for any of the three bodies of water in the area.
- The libraries data set is incomplete in that it does not include the library at Flagg Street School. In the map below, one can see the library (represented by an asterisk in the middle of the purple school point) at Notre Dame Academy on the right side of the map. The school farthest to the left, Flagg Street School, also contains a library but the library data set did not identify it. In addition, the libraries are point data and therefore do not show the exact location or shape of the libraries. However, the attribute information for the St. Justina Daley Library at the Notre Dame Academy is complete.

- Schools data set: The schools data set identifies all three schools in the area but it does not cover the entire school territories, as one can see in the above map. The aerial photograph of the three identified schools shows grey structures that are much larger than the point data provided by the MassGIS schools data. While all relevant features are present, the attribute information is incomplete for two of the schools, as fax numbers are not provided for Solomon Schechter Day School or Notre Dame Academy, and there is no street address provided for Solomon Schechter Day School.
- Prime Forest data set: The prime forest data set has complete attribute data and all relevant features (forest area) are present but it does not cover all of the forested areas in the neighborhood. In the map above one can see that the boundaries of the prime forest data (shown in light green) do not match the tree lines of the aerial photograph.
- Land Use data set: The land use data set is very complete as all of the attribute information is complete for all features, it covers all of the study area, and the relevant features (forest, water, high density and medium density residential areas, and urban public areas) are present.
8) The metadata for the various data sets provides information regarding the time period for which the data is relevant. Many of the data sets were created at different times and therefore have different up-to-date statuses. The data seems to range in relevance from 2000 to 2007.

- The MassGIS road centerlines data is up to date as of December of 2007, whereas the StreetMap USA road centerlines data is up to date as of 2000 and the Census TIGER road centerlines data is relevant as of 2000. The MassGIS hydrography data set is up to date as of October of 2004, and the Census TIGER hydrography data is relevant as of 2000. The MassGIS school data set is relevant as of December 28, 2007, the libraries data set as of October 2004, the Prime Forest data set as of 2007, and the land use data set as of January of 2002.

9) From a qualitative perspective, the attribute accuracy of the data sets ranges from poor, as in the Census TIGER hydrography data set, to very good, as in the MassGIS land use data set.

- Out of all three road centerlines data sets, the MassGIS road data is the most adequate for the project needs, as it provides the most street name data as well as the most extensive coverage of the area. Because the StreetMap USA and the Census TIGER road centerlines do not have data for many local roads, they are inadequate data sources for this project, as most roads around the schools are small local roads.
- The Census TIGER hydrography data is inaccurate in its attribute information as it does not have data for any of the bodies of water in the area. The MassGIS hydrography data is more accurate, although it does not identify all bodies of water. As stated in Question 3 (displayed in the 2\textsuperscript{nd} map), there is an unidentified stream next to one of the schools. This lack of attribute information could lead walkers right into the stream.
- School data: While Notre Dame Academy is properly attributed, Flagg Street School and Solomon Schechter Day School have inaccurate attribute information. At Flagg Street School, the Principal listed, Gerald Hippert, has retired. The new principal, Sheila Graham, is not listed. In addition, the school grades listed are Kindergarten through sixth grade, but this school also has a preschool. The Solomon Schechter Day School attribute information is also incorrect (as seen in the image below) as Solomon Schechter Day closed about three years ago and a new school is now in its place. The new school is called the Summit Academy School for Alternative Learners. The metadata for the schools data set says that the data was updated on December 28, 2007. However, Solomon Schechter closed before December 28, 2007, which leads me to believe that this data is not well-maintained.
• While the attribute data for the one library displayed in the Library data set is correct, the library data set does not account for the library in Flagg Street School. The metadata suggests that this may be because the library was “not registered with the Massachusetts Board of Library Commissioners (MBLC) as of October, 2004.”
• While the Prime Forest data does not cover all of the forest territory in the area, it is adequate for this project because it identifies nearly all of the forest and covers nearly all of the forest territory. In addition, the attribute information (acres, shape, and type of forest) is extremely complete.
• The land use data set is very useful in this project as the attribute information is quite accurate. The land use codes (forest, water, high density residential, medium density residential, and urban public areas) are very accurate, and thus very valuable, in helping to determine the best paths to the schools for walkers and drivers.