RESEARCH & DEVELOPMENT

Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data

Timothy Mulrooney, PhD
Department of Environmental, Earth and Geospatial Sciences
North Carolina Central University

NCDOT Project 2014-02

FHWA/NC/2014-02

February 2015

FINAL REPORT

Applying QA/QC Procedures to Quantitatively Measure the Quality of NC DOT GIS Data

North Carolina Department of Transportation Research Grant 2014-02
In Support of Report Number FHWA/NC/2014-02

Submitted by

Timothy Mulrooney
Department of Environmental, Earth and Geospatial Sciences
North Carolina Central University

Technical Report Documentation Page

1. Report No. FHWA/NC/2014-02	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data		5. Report Date 2/20/2015
		6. Performing Organization Code
7. Author(s) Timothy Mulrooney		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Environmental, Earth and Geospatial Sciences 2202 Mary Townes Science Complex 1800 N. Fayetteville Street Durham, NC, 27707		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address North Carolina Department of Transportation Research and Analysis Group 1 South Wilmington Street Raleigh, North Carolina 27601		13. Type of Report and Period Covered Final Report 8/15/2013 – 8/15/2014
		14. Sponsoring Agency Code 2014-02
Supplementary Notes:		

16. Abstract

The current state of the NCDOT Roads GIS database is unknown. While adequate metadata does exist to explain various descriptive, structural and administrative information about the data, it is very general at best and does not include specific revisions and updates that have occurred throughout the life of the data. As a result, a need exists to explore various dimensions of data quality within this robust database. While verifying each of the individual features and accompanying attributes is an impossibility within the scope of this research project, this research looked to explore various QA/QC (Quality Assurance/Quality Control) assessment techniques and integrated procedures within the framework of accepted QA/QC standards. This goal of this research project was to explore the integrity of the NCDOT GIS database at various scales. A result of this research was to give the NCDOT a snapshot of their database and if it is sub-standard, a direction to dictate data development or re-development efforts as part of later in-house or contract work.

17. Key Words Geographic Information Systems, Quality Assurance, Quality Control, Mapping, Data Quality Mapping, GIS Editing		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 63	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

DISCLAIMER

Acknowledgements

Dr. Timothy Mulrooney and the NCCU Research Team wishes to thank the NC DOT for the opportunity to work on a project of this magnitude. This project has provided valuable work experience for students getting ready to enter the workforce. NCCU is also grateful for input provided throughout the course of this project. While not fully complete, those who helped facilitate this project included: Jonathan Arnold, Ashley Clowes, Tim Sheldon, Faith Johnson and Mustansir Kadibhai. Timothy Mulrooney and NCCU are happy to follow-up on this work or assist those who have questions about this document. His contact information is below:

Timothy Mulrooney
Department of Environmental, Earth and Geospatial Sciences
North Carolina Central University
1246 Mary M. Townes Science Complex
1801 Fayetteville St., Durham, NC, 27707
Phone: 919-530-6575
E-mail: tmulroon@nccu.edu

Table of Contents

1. Introduction

The North Carolina Department of Transportation (NC DOT) maintains comprehensive databases of their spatial assets and related features within a GIS (Geographic Information System). It is published on a quarterly basis. There are three separate roads features maintained by the NC DOT. The **LRS_Arcs** feature class contains individual named road segments for the entire state and more than 205,000 **LRS_Arcs** are contained within the state. The NC DOT also maintains a **Road Characteristics** database which contains attribute information such as the number of lanes, speed limit, median type, median width about each of the aforementioned named roads in the **LRS_Arcs** database. A new Road Characteristic segment is required when a particular attribute changes (number of lanes changes from 2 to 4, for example) along the course of an **LRS_Arcs** segment. As a result, there are more than 400,000 segments that compose the Road Characteristics database, each of which can be linked to an **LRS_Arc** segment using a unique identifier called a ***key***. An example of a separate Road Characteristic can be seen in the diagram above for the same **LRS_Arc** where a Road Characteristic along the **LRS_Arc** (in this case *STRCTR_CD* which shows the presence of a bridge) has changed.

2. Goals

Both the spatial and attribute information that compose this database were created using assorted techniques at different times by various personnel over the past 30 years. The quality of the NC DOT Roads GIS data is unknown. While adequate metadata does exist to explain various descriptive, structural and administrative information about the data, it is very general at best and does not include specific revisions and updates that have occurred throughout the life of the data. As a result, a need exists to explore various dimensions of data quality within this robust database. While verifying each of the individual features and accompanying attributes was an impossibility within the scope of this research project, this research looked to explore various QA/QC (Quality Assurance/Quality Control) assessment techniques and integrated procedures within the framework of accepted QA/QC standards. A result of this research was to give the NC DOT a snapshot of their database and if it is sub-standard, a direction to dictate data development or re-development efforts as part of later in-house or contract work.

2.1 North Carolina Central University

The Department of Environmental, Earth and Geospatial Sciences (DEEGS) at North Carolina Central University (NCCU) performed the work for this project. The department began preliminary work in August, 2013, with much of the work being performed during the Summer of 2014. The DEEGS offers undergraduate programs in Environmental Science, Geography and GIS, as well as a graduate degree in Earth Science. The department's mission is to promote intellectual, professional, and personal excellence through the highest quality instruction, research, and service. Its vision is to be recognized as a regional, statewide, and national resource for students and society as well as professionals who work in the many fields that are encompassed by the environmental, earth, and geospatial sciences. The careers goals of recent DEEGS graduates has been a healthy combination of public sector professional work (EPA, State Agencies), private contractor work and the pursuit of Master's and Ph.D. degrees at NCCU (Master's only) and other institutions.

2.2 The Research Team

Timothy Mulrooney is an Assistant Professor in the Department of Environmental, Earth and Geospatial Sciences (DEEGS) at North Carolina Central University (NCCU). The focus of his teaching and research is GIS and the application of GIS to a variety of disciplines that NCCU offers. Before his tenure at NCCU, he worked as a Senior GIS Analyst with the Army SRP (Sustainable Range Program) GIS Regional Support where he provided GIS analysis, support and database administration for Army assets throughout the world. In this research project, he served as the Principal Investigator and managed every aspect of this project, developed Python code and developed the reporting procedures for the project.

Glenn Koch is a 2^{nd} year graduate student in the DEEGS at NCCU. He served as a technical lead on this project and helped manage databases used in the course of the project. He kept track of individual progress through each phase of the project, developed Python code and models to streamline tedious procedures. Where needed, he performed QA/QC on the database to check for attribute and geometric accuracy.

Corey Koch, Edward Holley and Roderick Mitchell also served on the research team and performed QA/QC work to check data integrity of the county-level databases. Corey is an undergraduate History Major (Anthropology Minor) at Virginia Commonwealth University in Richmond, VA. He was recruited onto this project because of his prior work on a GIS internship in Elizabeth City, NC, in 2013. Edward Holley is an undergraduate student at NCCU. He was a student of Tim Mulrooney in the GIS courses that he taught at NCCU. Roderick Mitchell is a May, 2014, graduate of NCCU. His degree is in Geography with a concentration in GIS. He was also a student of Tim Mulrooney during his time at NCCU. Both Edward and Roderick have been exemplary students and are well respected by all DEEGS faculty. Roderick was hired by a utilities company in Raleigh soon after this project ended. Edward will be graduating in May, 2015.

(left to right) Front Row: Corey Koch, Tim Mulrooney
Back Row: Roderick Mitchell, Edward Holley, Glenn Koch

3. Components of Data Quality

Various components contribute to GIS data quality. Within this study, horizontal accuracy, attribute accuracy and attribute completeness were checked in the database. For the horizontal accuracy, the number of **LRS_Arcs** to be checked are based on the total number of **LRS_Arcs** within each county. For Alamance County, there are 2,343 **LRS_Arcs** within the county. Based on the ANSI (American National Standards Institute) Z1.4-1993 Standard, the number of random features originally selected to be checked for accuracy within this county is 125.

It must be noted that for each **LRS_Arcs** database, not all features were candidates for random selection. Only those features under the purview of the NC DOT were selected. These include features attributed as *Interstate*, *US Route*, *NC Route* and *Secondary Road* through the RT1_CLSS_CD attribute field. Those features not eligible for selection include *Ramps*, *Park Roads* and *Local Roads*.

While a length criterion for the **LRS_Arcs** was not stipulated for this project, the NC DOT works on projects in which a minimum **LRS_Arc** length must be satisfied. A script was produced which measures and catalogues the length of random samples originally selected for each county. In some counties, approximately 10% of **LRS_Arcs** randomly selected have a length of less than .1 mile (528 feet). These counties include Alleghany, Avery Duplin and Hertford Counties. Counties with more than 40% of **LRS_Arcs** less than .1 miles in length include Wake, Cumberland and New Hanover County. A general trend shows that more urban counties have a higher percentage of, and therefore shorter, **LRS_Arcs**. The result of this analysis is highlighted in Table 1, Appendix B.

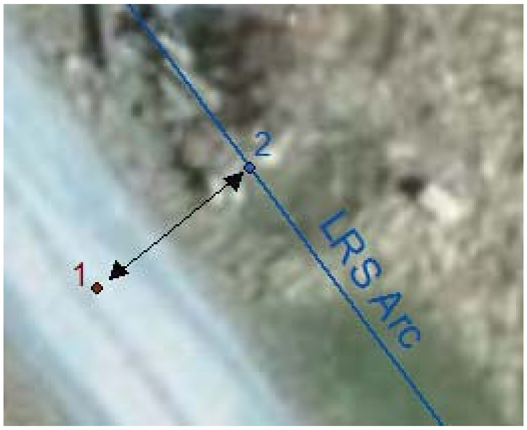
In Alamance County, 125 **LRS_Arcs** are related to 282 Road Characteristics which compose the entire length or portion of these 125 **LRS_Arcs**. In many cases, the **LRS_Arc** and accompanying Road Characteristic exhibit a 1:1 relationship, meaning that the **LRS_Arc** maintains the same Road Characteristics for the entire length of the **LRS_Arc**. In fewer cases, a 1:MANY relationship will be exhibited between the **LRS_Arc** and the Road Characteristic, all of which must be checked by hand. In the database development code (highlighted in Code Sample 1, Appendix C), a relationship class between the **LRS_Arcs** and Road Characteristics using the G1_FtSeg field as a key was created. In this way, users can link selected **LRS_Arcs** and the accompanying Road Characteristic(s) propagated both forwards (from selected **LRS_Arcs** to Road Characteristics) and backwards (selected Road Characteristics to **LRS_Arcs**).

3.1 Horizontal Accuracy

Horizontal accuracy represents the distance that a GIS data layer deviates from geographic reality. It essentially measures the distance a GIS data feature is from where it 'should' be. While is impossible to tell the exact location of where a feature should be placed, as geo-rectified imagery and even high precision GPS have inherent error attached to them, imagery used to confirm horizontal accuracy was the latest DOQ (Digital Ortho Quadrangle) imagery provided by NC One Map available through the NC One Map service (http://services.nconemap.com/arcgis/rest/services).

Given that the error for each individual feature varies as one travels along that feature, a uniform method was determined to measure this accuracy. In consultation with the NC DOT team, horizontal accuracy was measured from the randomly selected feature to the road centerline as portrayed on the imagery. Along each **LRS_Arc**, three equally spaced points (one in the middle, two others halfway between the middle and each end) were created by

NCCU staff and the distance was measured from each of these three points to the centerline at a 90º angle and measured using the *Near* Function. In the diagram to the left, Point #2 is located along the **LRS_Arc** at its midpoint. Point #1 was created by NCCU in an edit session and the *Near* function measures the distance between Point #1 and Point #2.



Since the three equidistant points are linked to the originating **LRS_Arc** in the *Near* function, statistics about each individual feature can be calculated. These tables can be *Summarized* and *Joined* back to the original **LRS_Arc** to determine which feature has the lowest average horizontal error within each county. Furthermore, all features within each county can be *Summarized* to find the following statistics about each measured error: Minimum, Maximum, Range, Average and Standard Deviation. These county and district-level maps are highlighted in the results and Appendix A.

The process of creating the 3 points to be compared to the digitized centerline points are shown in Code Sample 2 in Appendix C. While digitizing points along the centerline needs to be done by hand, running the *Near* function and *Summarizing* the data were run afterwards using the Python script in Code Sample 3 in Appendix C.

3.2 Attribute Accuracy

Attributes are the non-spatial characteristics of an entity. NC DOT attributes are uniform across an entity, and serve to distinguish one object from another. Attribute values can be unstructured text descriptions (e.g., Street Name = 'SR-2392') or numerical values (Surface Width = 32). In other cases the NC DOT uses domain fields to store values that can be described using a domain table. For example, the Right Turn Lane Type fields can only have one of four different values: 1 = MULTI_TURN_LANE_OR_BAYS , 2 = CONTINUOUS_TURN_LANE, 3 = SINGLE_TURN_BAY and 4 = NO_TURN_DURING_PEAK_PERIOD. Values of "Null" for this field mean that there is no right turn lane present for this section of road. These Road Characteristics need to be confirmed using a veritable source, either from imagery or existing documentation.

Each **LRS_Arc** will be assigned a percentage score based on the total length of the **LRS_Arc** that is correct based on the related Road Characteristics. For example, an **LRS_Arc** has a length of 1000 feet, with one Road Characteristic segment having a length of 400 feet (RC #1) and the other having a length of 600 feet (RC #2). If the surface width for RC#1 has an incorrect value for its entire length, but RC#2 is attributed correctly, this score for that particular attribute will be 60. Other attribute scores for the same **LRS_Arc** may earn different scores based on the percentage of Road Characteristics that are correct. Attribute Accuracy will return a score between 0 and 100, representing the percent of **LRS_Arc** length that has been correctly attributed. They will be agglomerated at the county level for each attribute checked.

In many cases, a Road Characteristic (speed limit, surface width, number of lanes) is applicable for all **LRS_Arcs**. However, in other cases, only a certain number of **LRS_Arcs** will contain a particular feature. In Alamance County, only 10 **LRS_Arcs** contain a left turn lane. The other 115 **LRS_Arcs** returned Null values for a left turn lane Road Characteristic, meaning that they correctly did not and should not have had (as confirmed by the NCCU Research Team) a turn lane associated with it. It is important to check data integrity for only these 10 **LRS_Arcs** that have left turn lanes. In cases of Turn Lane Type/Length, Structure Type and Median Type/Width, 1 attribute value will be returned, representing the percent of correct attribution for *only those features present in the **LRS_Arcs** database*. This will give a better refection of attribute accuracy for the NC DOT and be used in measuring attribute accuracy and overall score calculation.

Table 2 in Appendix B highlights the attributes to be checked for each **LRS_Arc** and its associated Road Characteristics. In some cases, dependencies exist between variables to be checked, which make verification easier. For example if a road does not have a median, then it will not have a median width. Road attributes were hand-checked from the following sources to ensure they agreed with a baseline standard: NC One Map Image Server, TEAAS Database, Google Maps and Streetview.

3.3 Attribute Completeness

Attribute accuracy looks to determine if the correct values have been populated for required attributes. It is impossible to check every single attribute within the database, as NCCU staff will have checked a sample of features within each county. However, we can ensure that all required attributes are actually populated and can apply rules as dictated by the ARID documentation provided by the NC DOT. Since the Road Characteristics database contains the actual information about each road segment, every Road Characteristic segment within the county (not just the random **LRS_Arc** or related Road Characteristic) will be checked for attribute completeness.

Attribute completeness measures the degree to which all required attributes have been populated according to the rules dictated by the NC DOT. This does not necessarily mean that they are correct. For example, the shoulder type must be populated and can be one of only six possible values matched through a domain table: 1 = GRASS_OR_SOD, 2 = GRAVEL_OR_STONE, 3 = BITUMINOUS, 4 = CURB_BITUMINOUS, 5 = CONCRETE, 6 = CURB_CONCRETE. For the median type, it can be only one of seven possible values if it is present. Otherwise, the value for this field is Null if no median is present. Free text or numbers must be populated in other fields. For the surface width field, a positive number must be present in order for it to be correct. Null is not an acceptable value. For the width of a turn lane, a positive value must be present if a turn lane exists for that portion of the road. In this case, Null is an acceptable value if no turn lane is present. These 34 ***attribute requirements*** are highlighted in Table 3 in Appendix B.

The ARID document supplied by the NC DOT team also highlights ***dependency requirements*** which are required for the adequate population of attributes. For example, if a median type is specified, then the median width must be populated. Conversely, if the median width field has been populated, then the median type field must also be populated. These 8 dependency requirements are highlighted in Table 4 in Appendix B.

A program using the Python programming language was written and a report generated to inspect all required attributes within the roads database based on certain domain fields or required values. In addition, an error log highlighting all errors was generated using the G1_FtSeg field to identify individual Road Characteristics that violated attribute and dependency specifications. Both the attribute requirements and dependency requirements will be catalogued, with a combination of both accounting for the attribute completeness score as part of this report card.

4. Literature Review

GIS data quality is the end-product of processes designed to ensure that newly created data are correct (Quality Assurance) while also identifying existing data that are incorrect (Quality Control). Applications of QA/QC extend well beyond the GIS world, such as banking, manufacturing, software, medicine and even taxonomy (Chapman, 2005). While Taulbee (1996), among others, distinguish between QA and QC, they are usually termed as a pair and feels that one can not exist independent of the other. The subject of QA/QC with respect to data accuracy is typically regarded as a business process in the GIS world, so there is relatively little academic literature focusing on GIS QA/QC theory. Instead, various organizations have documentation and processes in place to define database schema and ensure that the various accuracies are adhered to that best fit their needs, resources and limitations.

Early pioneers of GIS recognized the importance of data quality, not only from a cost efficiency standpoint, but because of the legal ramifications in publishing incorrect spatial information which may lead to accidents or the incorrect use of the data (Epstein, 1988). Even then, they understood the compromise between accuracy, the cost of creating accurate data and the inevitability that some error will still exist. It is unreasonable to expect the NC DOT to photo-revise every single road in the GIS database, re-attribute it correctly and then verify them using another party in a timely manner given current personnel and financial constraints. This compromise is what Bédard (1987) called *uncertainty absorption*. Regardless of resource allocation, verification of data quality should be done by discipline experts with a long term goal of developing data quality standards. This helps to protect the GIS data producer from the potential misuse of GIS data (Aronoff, 1989).

Given its nature, dimensions of geospatial data quality are difficult to assess and quantify. Before the large scale democratization of GIS data and technologies, Openshaw (1989), Buttenfield (1993) and Caspary and Scheuring (1993) recognized the importance of GIS data quality. Some research has gone into quality assessment within popular GIS journals. Devillers et al. (2007) and Howard (1996) explore data quality with a goal of creating tools to assess data quality and automate this process.

Metadata has been used to describe data quality measures taken during the data development process and subsequent updates. Most generally thought of as "data about data", metadata serves as a formal framework to catalog the lifeline of a particular GIS data set. More recently, feature level metadata (Devillers et al., 2005) has been able to capture data quality information, but is typically limited to quantitative measures of positional accuracy and qualitative information related to data lineage within eight of the more than 400 entries that comprise a complete FGDC (Federal Geographic Data Committee) compliant metadata file. Even now, the population of these metadata elements is not fully automated and given the fickle nature of the human component, data quality assessment via the extraction of metadata entries is not always the best approach.

With the permanent marriage of GIS to popular surveys such as the Decennial Census and American Community Survey (ACS) in order to map metrics related to housing, demographics, employment, socio-economics and education, data accuracy should be a concern for anyone wishing to utilize this type of information. This is even more of a pressing concern beginning with the 2010 Decennial Census. Detailed socioeconomic, employment and education data offered through the SF-3 (Short Form) and SF-4 forms below the county level are no longer available. These data are now only available through the ACS, which collects data more frequently and over shorter intervals (1, 3 and 5 year), but with a higher margin of error due to smaller sample sizes (MacDonald, 2006).

Given its quantitative nature, horizontal accuracy is probably the easiest to measure. Mapping census data has been made increasingly simple with the integration of TIGER/Line files into GIS systems to map states, counties and sub-county units such as census tracts, block groups and even blocks that can be linked to tabular data and visualized on a map (Broome & Meixler, 1990). Efforts are made to provide users with an adequate *scale denominator* via metadata to quantitatively generalize the horizontal accuracy and the processes used to create the data for display on maps. The National Mapping Accuracy Standards (USGS, 1947) and later National Standard for Spatial Data Accuracy (FGDC, 1998) dictate standards for horizontal accuracy for various map scales and the digital data that compose the maps.

Of more concern within the GIS community is quantifying attribute accuracy. For a relatively small project of this nature, the NC DOT required attribute assessment for 23 high-priority attributes (out of more than 90 NC DOT attributes) across more than 205,000 LRS_Arcs and 400,000 related road characteristic segments. Given sheer data volume, it is imperative to check if information is correct upon data creation compared to years after the data have been created. In most contemporary literature on the subject, attribute accuracy revolves around visualizing margins of error (MOE) for sampled data. The United States Census Bureau (2008) provides a clear understanding of the processes used in which these data are created and caveats for mapping these data. Xiao et al. (2007) and Wong and Sun (2013) address this specifically for the ACS. However, Wong and Sun assert that the mapping of attribute accuracy has not been fully and efficiently dealt with by the GIS community.

As with attribute accuracy, attribute completeness remains problematic to quantify and map. Considered an early pillar of geospatial error (Sinton, 1978), completeness, or the "extent to which information is comprehensive" (MacEachern, 2005; pp. 146) can take on different meanings if it refers to field collection, imagery or thematic classification (MacEachren 2005). Wong and Sun (2013) once again address the attribute completeness issue within the confines of the ACS, as some 1-year and 3-year estimates contain no data for a considerable number of enumeration units. However, this is more a byproduct of methods and sample size than the processes that may compromise data completeness. Like attribute accuracy, quantitative studies on the spatial variability of attribute completeness are still few and far between.

Agglomerating and mapping these three dimensions (horizontal accuracy, attribute accuracy and attribute completeness) of data quality for NC DOT GIS data serve as the cornerstone of this research. More of the literature on the mapping of data quality focuses more on the cartographic representation of data uncertainty (Leitner & Buttenfield, 2000; MacEachren, 1992) than feature-level data quality measures (Zandbergen, 2008). Other research has explored the notion that spatial data quality is spatially auto-correlated (Bierkens & Burrough, 1993; Smith et al., 2003; Foody, 2005). In other words, higher quality data will be located near other high quality GIS data, while lower quality data is located near other low quality GIS data. This research will assess and visualize this notion at the county level for the state of North Carolina.

5. Results

5.1 Horizontal Accuracy

Three points were created along each random **LRS_Arc** using custom Python code and the NCCU Research Team digitized three corresponding points along the centerline of the accompanying road from NC DOT imagery. In places where a viable **LRS_Arc** did not exist according to the imagery (as shown in the diagram to the left), the random **LRS_Arc** was not included in fear of skewing the overall horizontal accuracy of the county. There were seven instances where an **LRS_Arc** existed in the GIS data, but the imagery showed that a road did not exist in the proximity of the **LRS_Arc**. These may be cases where data were imported from the universe years ago, but no longer exist and have not been updated in the interim.

The *Near* function was used to compute the distance between each point along the **LRS_Arc** and the point on the digitized centerline. Therefore, horizontal accuracy was computed on a feature by feature basis for all randomly selected **LRS_Arcs** within the county and retained. Statistics for all **LRS_Arcs** were computed at the county level using the *Summarize* tool.

Average values for horizontal accuracy ranged from .5944 feet (Davidson County) to 18.8212 feet (Clay County). Other counties whose horizontal accuracy was less than a foot included Davie (.6185 feet), Stokes (.7437 feet), Cabarrus (.8824 feet), Forsyth (.8894 feet), Lincoln (.9248 feet) and Scotland (.9825 feet) Counties. Besides Clay County, other counties whose average horizontal accuracy were more than 10 feet included Hyde (10.1741 feet), Macon (10.2501 feet), Sampson (11.4437 feet) and Craven (16.2905 feet) Counties. A map for the state of North Carolina can be seen in Map 1 in Appendix A.

Standard deviation is a general descriptive metric that measures how far each **LRS_Arc** within each county deviates from the mean. While two counties may have the same average horizontal accuracy, a county with a higher standard deviation has more variation in accuracy from one measured **LRS_Arc** to the next. It is understood that counties with lower horizontal accuracies will have lower standard deviation, but these metrics can be graphed on a plot with a relatively high coefficient of determination (r^2) and outliers can be found. Values with the highest positive residuals (standard deviations which are relatively high for its accompanying horizontal accuracy based on a line of best fit) include Macon, Burke, Green, Durham and Avery Counties. Counties with the lowest negative residuals (where standard deviation is best based on the horizontal accuracy) include Craven, Perquimans, Johnston, Richmond and Chowan Counties. These are shown in Map 2 in Appendix A and summarized in Table 5 in Appendix B.

The NC DOT also maintains fourteen transportation divisions across the state. Using the *Dissolve* tool, horizontal accuracy and standard deviation was agglomerated at the division level and mapped as shown in Maps 4 and 5 in Appendix A. Divisions 9, 10 and 12 running along the Interstate Route 85 and US Route 52 corridors ranked the best with horizontal accuracies of .7702, 1.3323 and 2.1576 feet respectively. The poorest horizontal accuracy was found in Division 2 (eastern part of state) and Division 14 (extreme western part of state) at 7.3223 and 7.0388 feet, respectively.

5.2 Attribute Accuracy

A select number of attributes for each county were checked for attribute accuracy to determine if they matched with up the Road Characteristics provided by the NC DOT. 23 separate Road Characteristics were checked for attribute accuracy.

In some cases, a Python programming script was written to determine if the attributes from the **LRS_Arcs** and the Road Characteristics matched up each other as per the ARID documentation. This was done for the following attributes:

- STREET_NAME
- LOC_CNTY_CD: accuracy derived from **LRS_Arcs** layer clipped at the county level

In other cases, attribute accuracy was checked against a separate GIS data layer provided by the NC DOT and compared to the Road Characteristics and related **LRS_Arcs**. This was done for the following attributes:

- NHS_TYP_CD
- TRCK_RTE_TYP_CD
- SHN_TYP_CD
- RW_WID

In one case, **LRS_Arcs** were hand-checked against Road Characteristic information provided by the NC DOT. There were very few NC DOT-controlled roads for this layer and all counties earned values of 100 for this attribute:

- MLTRY_BASE_CD

In one case, the online layer provided by the NC DOT via TEAAS was used as the basis for Road Characteristics, thereby matching each **LRS_Arc** exactly. When attempting to confirm using them using road ordnances, limited inspection showed they also matched up correctly for the following attribute:

- SPD_LMT_TYP_CD

The following layers were checked against confirming imagery by the NCCU Research Team. In some cases, a particular attribute was not present for all **LRS_Arcs** and only those attributes where the feature was present or should have been present was computed for accuracy based on the percent of **LRS_Arc** length that was correct. This occurred for the following attributes:

- TRNLN_LFT_TYP_CD
- MDN_TYP_CD
- MDN_WID
- TRNLN_RGT_TYP_CD
- TRNLN_LFT_WID
- TRNLN_RGT_WID

The following attributes were confirmed from imagery by the NCCU Research Team and occurred across all **LRS_Arcs**. These attributes were:

- ACS_CTNRL_TYP_CD
- NBR_LANE_QTY
- SHLDR_LFT_TYP_CD
- SHLDR_RGT_TYP_CD
- SRFC_TYP_CD
- SRFC_WID
- SW_PVD_LFT_QTY
- SW_PVD_RGT_QTY

The following attribute was checked using GIS Data provided by the NC DOT to show the presence of the feature and confirmed using imagery to inspect its accuracy. This attribute did not occur for all **LRS_Arcs** and their frequency by county varied.

- STRCTR_CD (https://connect.NC DOT.gov/resources/gis/pages/gis-data-layers.aspx)

Overall, Attribute Accuracy values ranged from 72.4313% (Hertford County) to 95.7096% (New Hanover County) as shown in Map 6 and Table 6. As per the request of the NC DOT, other maps showing individual attributes such as Median Types, Turn Lane Types (Left and Right), Number of Lanes and Structure Type are shown in Maps 6a through 6e.

In terms of individual attributes, each county's percentage (percent of attribute length marked as correct) was averaged to determine the attribute that was least and most attributed correctly. RW_WID (49.23%) was most incorrectly attributed, followed by STRCTR_CD (51.27%), TRNLN_RGT_TYP_CD (68.20%), MDN_WID (68.29%) and MDN_TYP_CD (71.70%). For the attribute RW_WID, as well as TRNLN_RGT_TYP_CD and MDN_WID and MDN_TYP_CD, most of the incorrect values occurred because Null values were populated in the Road Characteristic when there should have been a value. In the case of STRCTR_CD, the appropriate **LRS_Arc** containing the structure did not have a separate Road Characteristic for the section of **LRS_Arc** that contained the structure.

It must be noted that some features were not present along the randomly selected **LRS_Arcs** and that some counties had absolutely no attributes for a particular attribute. For example, 10 counties (Alleghany, Madison, Mitchell, Moore, Rutherford, Sampson, Tyrrell, Vance, Warren and Yancey) did not have a left turn lane from the randomly selected **LRS_Arcs** and were assigned an attribute accuracy score of 100%. 23 counties did not have a right turn lane and 2 counties did not have a median. All counties had at least 1 structure. It must be noted that attributes with small sample sizes may be the victim of the *small number problem* (i.e. the only feature that was measured happen to be the only incorrect one) and should be treated accordingly. The number of attributes appearing in each county from the randomly selected **LRS_Arcs** has been provided in the final summary table by the PI. A breakdown of all attributes for all counties has been provided in digital format by the PI and a summary of this breakdown is shown in Table 7 of this document.

5.3 Attribute Completeness

A Python programming script was run to determine if attributes within the Road Characteristics database adhered to domain and/or population requirements. Upon further inspection of the data and ARID documentation provided by the ND DOT team, attribute completeness entails both the population of required attributes and while checking to ensure that dependencies exist between certain attributes. All 409,377 eligible individual Road Characteristic arcs (not just the randomly selected **LRS_Arcs**) were checked for the 34 attributes and 8 dependencies as highlighted in Tables 3 and 4 in Appendix B. As a result, almost 14 million attributes where checked for completeness and more than 7 million attributes were checked as part of the dependency requirements stipulated by the ARID documentation.

In terms of the 34 attributes, 8 different attribute errors were detected and 2.1935% of attributes were incompletely populated. The most common attribute population error was the absence of a right of way (Error #14 where 34.5133% of attributes were not populated correctly), followed by left shoulder type (Error #15 where 14.177% of all attributes were incorrectly attributed) and right shoulder type (Error #16 where 13.08% of all attributes were incorrectly attributed). Other attribution errors found were incorrect speed limit (Error #17 at 4.909%), incorrect surface width (Error #19 at 3.1633%), incorrect surface type (Error #18 at 2.9472%), incorrect number of lanes (Error #12 at 1.7849%) and the incorrect median type (Error #9 at .0002%). In the case of shoulder types (Errors #12 and #13), speed limit (Error #17) and surface type (Error #18), only certain domain values are allowed and Null values are not allowed. In the case of surface width (Error #19), a positive non-Null number is required. In the case of median type (Error #9) a domain value or a Null value is allowed. In that case, an illegal domain value was populated.

Mecklenburg County led all counties with 3.8697% of all attributes incompletely populated, followed by Vance (3.3390%), Chatham (3.1459%), Guilford (3.1326%) and Haywood (2.9755%) Counties. Dare County had 1.2496% of all attributes incompletely populated, followed by Pamlico (1.3335%), Lincoln (1.3391%), Currituck (1.3442%) and Alexander (1.4352%) Counties. Table 8 in Appendix B highlights this information and Map 8 in Appendix A shows the county totals for attribute population.

In terms of dependency errors, 24.985% of all possible dependencies earned incomplete values and all dependency errors were flagged at least once. The most common dependency error was Error #38 (86.85%), which required that all four improvement attributes be populated or all have values of Null. Even if three of the improvement attributes were populated, it would be marked as incorrect if it were just missing one attribute. Other common errors revolved around the addition date entries and their co-population with addition document type (Error #39) and addition document ID (Error #41). In many cases, it appeared that a legal, but default value, was placed in there. If the date were after 12/31/1930, these errors would be flagged if the accompanying attributes were not populated. Since more than 27% of all attributes were flagged for Error #42 (addition date is before then improvement date), this may be case of legal, but erroneous values being populated for the addition date. The least frequent dependency errors flagged was Error #40 (addition document ID and addition document type are not co-populated) at .0144%, followed by Error #35 (median type and median width are not co-populated) at .1742%, Error #36 (left turn lane type and width are not co-populated) at .5789% and Error #37 (right turn lane type and width and not co-populated) at 1.1200%.

Finally, all 42 errors were agglomerated to determine attribute completeness as a product of attribute and dependency requirements. Overall, 6.53% of the entire dataset was incomplete based on these 34 attribute and 8 dependency requirements, with a brunt of the incorrect entries being dependency errors. Combined, the best counties were Pamlico (4.6283% incorrect), Anson (4.6915%), Greene (4.8680%), Sampson (5.1689%) and Stanly (5.1744%) Counties. On the other end, Mecklenburg County (8.4794%) has the most incomplete attribute data, followed by Vance (8.0819%), Guilford (7.7219%), Chatham (7.5955%) and Dare (7.4291%) Counties.

In all, more than 305,000 attribute and 818,000 dependency errors were found from the 42 possible errors among 409,377 individual Road Characteristics checked in this database. All 1,123,350 (out of a possible 17,193,834) errors were written to an error log, identifying each error by county name, G1FtSeg attribute, error # and a short description of the error. A copy of this file was given to the NC DOT for future use at the conclusion of this project.

5.4 Overall Accuracy

For each of the three categories (Horizontal Accuracy, Attribute Accuracy and Attribute Completeness) each county earned a score between 0 and 100. Scores were linearly scaled based on the value in question and its relationship to the range of values for each category. For example, Lee County earned an average Horizontal Accuracy of 4.5756 feet. If the minimum value was .5946 feet and the range of all Horizontal Accuracy values was 18.2266 feet, Lee County would earn a score of 78.2 out of 100. For Attribute Completeness, New Hanover County had 7.2856% of all attributes incomplete. Given the best county was 4.6283% and the poorest county was 8.4793%, New Hanover County would earn a score of 31.0. For each category, the highest score would earn a value of 100 while the lowest score, whether measurement (Horizontal Accuracy), percent correct (attribute accuracy) or percent incomplete (Attribute Completeness), would earn a score of 0. A final score was earned by averaging the 3 scores to yield a final score between 0 and 100.

The final scores ranged from 26.1 to 88.2 and are shown in Table 9 and Map 10. Stokes County was rated the best county based on scores of 99.2 (Horizontal Accuracy), 92.1 (Attribute Accuracy) and 73.3 (Attribute Completeness). Other highly rated counties were Union (84.5 Final Score), Rowan County (84.1), Stanly County (84.0) and Pamlico (83.1) Counties. The lowest rated counties were Craven County with a score of 26.1, followed by Dare County (39.3), Durham County (39.6), Chowan County (39.8) and Burke County (40.9). A standard deviation at the county level was run on all 3 metrics to determine counties whose values did not deviate very much. Henderson County, with values of 72.3 (HA), 72.7 (AA) and 75.2 (AC) had high quality data that did not vary much. Mecklenburg County, on the other hand, had values of 90.2 (HA), 82.7 (AA) and 0 (AC). Vance County also had high quality data in 2 categories, but poor attribute completeness.

5.5 Other Observations (Urban vs. Rural)

As per the request of the NC DOT, a breakdown of various attributes by urban and rural status as per the URBAN_ID_CD attribute in Road Characteristics was performed. A comprehensive summary is attached in the spreadsheet given to the NC DOT. **LRS_Arcs** were classified as 'URBAN' or 'RURAL' based on the related Road Characteristics. In looking at access control (ACS_CNTRL_TYP_CD), 98.50% of urban **LRS_Arcs** were correct while 99.28% of rural **LRS_Arcs** were correct. For the number of lanes

(NBR_LANE_QTY) attributes, there was no discernable difference between the 2 (urban = 99.83% vs. rural = 99.86%).

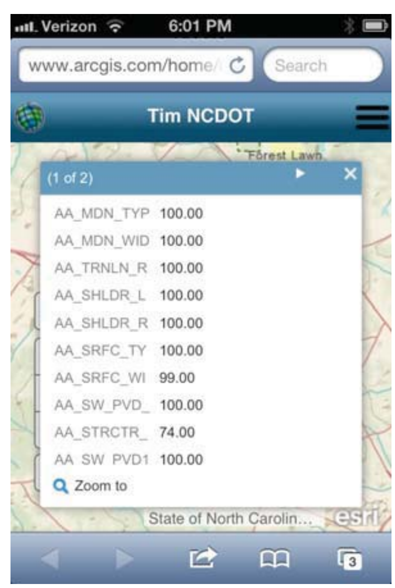
In addition, turn lanes and medians (type and width) were extracted, broken down and summarized at the county level based on the appropriate road characteristic. For Alamance County's 10 left turn lanes, 8 were classified as urban while the other 2 were classified as rural. Overall, of the 718 left turn lanes that occurred within the randomly selected **LRS_Arcs** candidates, 419 were classified as urban while the remaining 299 were classified as rural. There were much fewer right turn lanes (234 urban/155 rural) and many more medians (607 urban / 490 rural) among the randomly selected features in the 100 counties, although there were more rural **LRS_Arcs** from the randomly selected features.

Given the concentration of urban features within the state and the random selection of candidate **LRS_Arcs**, many counties did not have an equal distribution of urban and rural features which were assessed for horizontal and attribute accuracy. For example, Currituck County had absolutely no features classified as urban among the 17 left turn lanes, 4 medians and 6 right turn lanes among the candidate **LRS_Arcs**. On the other hand, Wake County only had 1 median classified as rural, with the other 15 left turn lanes, 7 right turn lanes and 41 medians begin classified as urban. Counties that did have an adequate number of features and a somewhat equal distribution between urban and rural **LRS_Arcs** included Brunswick, Lenoir, Nash and Richmond Counties.

Given the limited number and unequal distribution of features among counties, it was difficult to make urban to rural comparisons at the county level. However, when summarized for the entire dataset, discernable differences can be noted. For the left turn type, 82.876% of urban attributes were correct compared to 74.310% for rural **LRS_Arcs**. For the left turn lane width, this difference (84.69% urban vs. 79.94% rural) is not quite as noticeable. For the right turn lane type, urban **LRS_Arcs** (84.321%) ranked better than their rural counterparts (70.492%). For median type, urban **LRS_Arcs** outscored their rural counterparts 82.025% to 69.042%. These comparisons are highlighted in Table 10.

6. Ground Truthing

In places where logistically possible, ground truthing was performed on the completed **LRS_Arc** database to confirm attribute accuracy where street-level imagery (Google Street View) did not exist. While many attributes could be checked using updated imagery, other attributes such as surface type, shoulder type and speed limit can be confirmed using manual inspection on site. Time constraints prevented large-scale ground truthing of data, but the attributes from 50 **LRS_Arcs** were checked. In order to do this, a select number of features from the final **LRS_Arcs** database were exported to an arcgis.com project, which could be accessed by a hand-held tablet or phone in the field. Pertinent attribute information from the hand-held device (see diagram to right) was accessed and compared to the actual road to determine if the inspected data were correct. An equal number of **LRS_Arcs** that contained turn lanes, medians, structures and **LRS_Arcs** with no major features were selected for ground truthing.



Attribute checked with median (G1_FtSeg: 195126)

Attribute checked with turn lane (G1_FtSeg: 40432)

Attribute checked with no major features (G1_FtSeg: 202976)

In all, 49 of the 50 **LRS_Arcs** that were checked had correct attribution. In one case, a left turn lane and left turn lane width was assigned a value of Null for the entire length of the **LRS_Arc** from the Road Characteristics when a turn lane did exist. It was re-assigned a score of 95, based on the percent of road length that was correct. In another case, an **LRS_Arc** with a length of 16,190 feet had 14 Road Characteristics related to the one **LRS_Arc**. All but one of the Road Characteristic arcs had correct attribution. However, this Road Characteristic with incorrect attribute only had a length of 2.44 feet, giving it a score of 99.98, which was rounded up to 100.

7. Recommendations for the NC DOT

A major part of this project was to assess the integrity of the NC DOT GIS database and provide recommendations for future work. The following are based on our findings:

- Horizontal Accuracy (summarized in *Horizontal Accuracy* tab of spreadsheet)
    - Horizontal Accuracy for the following counties is over 8 feet and should be addressed when time allows (from poorest to better horizontal accuracy): Clay, Craven, Sampson, Macon, Hyde, Yadkin, Pender, Avery, Cherokee, Johnston, Tyrrell, Watauga, Beaufort, Greene and Gates. Other non-photo revised counties should be addressed in order, from poorest horizontal accuracy to best horizontal accuracy.
    - In graphing Horizontal Accuracy versus the standard deviations of measurements, a few counties had very high variability compared to their measurements. These counties may need to be re-visited at some point in time to 'tighten' up their data. These counties include: Durham, Macon, Greene and Burke.
- Attribute Accuracy (summarized in *Attribute Accuracy* tab in spreadsheet at the county and attribute level)
    - For the randomly selected features, average attribute accuracy for all counties was over 87%. Counties whose attributes were below 80% include (from least to greatest) the following counties: Hertford, Durham, Davidson, Greene, Avery, Forsyth, Lincoln, Dare, Cabarrus, Jones, Craven, Jones, Craven, Caswell, Harnett, Camden, Burke, Chowan, Caldwell and Brunswick.
    - Less than ½ (49.23%) of all Right of Way Width (RW_WID) attributes were not correct. In many cases, the actual values placed in by NC DOT were correct, but many of the accompanying Road Characteristics contained Null values. A place to start would be Caswell County, where 27% of the Right of Way attribute was correct.

- Barely more than ½ (51.27%) of the Structure Type attributes (STRCTR_CD) were correct. This was mainly due to the fact that the **LRS_Arc** that contained the structure did not have a separate Road Characteristic segments to signify the presence (or discontinuation if it were coded as having a structure) of the structure along that segment of road. The **LRS_Arc** needs a separate Road Characteristic segment to signify the presence of a structure. A place to start would be Scotland County, where less than 2% of the STRCTR_CD attribute was correct.
    - For the turn lanes types and widths (TRNLN_LFT_TYP_CD, TRNLN_RGT_TYP_CD, TRNLN_LFT_WID, TRNLN_RGT_WID), Road Characteristics contained Null values when there should have been a valid value. Very few were miscoded.
    - For the median type and width (MDN_TYP_CD and MDN_WID), Road Characteristics contained Null values where there should have been a valid value. Very few were miscoded.
    - In some cases, the street name (STREET_NAME) did not exactly match the Rte_Nm attribute in the **LRS_Arcs** table. These should be revisited to ensure the naming conventions agree with NC DOT standards.
    - Attribute accuracy for rural **LRS_Arcs** for the entire database was less than that for their urban counterparts for attributes related to turn lanes and medians. It is suggested that rural **LRS_Arcs** within each county be revisited when time and resources allow.
- Attribute Completeness (summarized in *Attribute Completeness* tab in spreadsheet)
    - An error report highlighting all 1,000,000+ attribute and dependency completeness errors has been provided in CSV format that can be viewed and sorted by county, G1_FtSeg and Error Number.
    - Mecklenburg County had with poorest attribute completeness, with almost 8.5% of all Road Characteristics incomplete as per ARID documentation followed by Vance (8.1%), Guilford (7.72), Chatham (7.60%) and Dare (7.43%) Counties.
    - 34 different attributes were checked for completeness.
        - For the Right of Way Width (RW_WID), 35.93% of values were Null. Note that this value from Attribute Completeness is higher because that value is weighted by **LRS_Arc** length and the number of features in the county for just randomly selected **LRS_Arcs**. A Null value is not acceptable for this attribute.
        - Both shoulder types (SHLDR_LFT_TYP_CD and SHLDR_RGT_TYP_CD) were incomplete for more than 13% of all Road Characteristics. A Null value is not acceptable for this attribute.
        - Almost 5% of all Speed Limit value (SPD_LMT_TYP_CD) were incompletely populated. This attribute must have 1 of about 20 different domain values and can not be Null.
        - Approximately 3% of Surface Type and Width (SRFC_TYP_CD and SRFC_WID) were incompletely populated. Surface type must be a coded domain and surface width must be a positive number. Neither can be Null.
        - The Number of Lanes (NBR_LANE_QTY) attribute must be populated with a positive number. Almost 2% of the values were populated as Null and incorrect.
        - The Median Type (MDN_TYP_CD) can be a coded domain or Null. However, in Lenoir County, it had an illegal domain value.

- 8 different attributes were checked to make sure they agreed with other attributes that needed to be concurrently populated
    - There are 4 different improvement attributes. All 4 must be populated if one has been populated. For almost 87% of Road Characteristics, that is not the case.
    - In .17% of all Road Characteristics, the Median Type (MDN_TYP_CD) and Median Width (MDN_WID) were not co-populated. If a Road Characteristic segment has a median type, then it must have a median width. If it has a median width, then it must have a median type.
    - The Turn Lane Type and Turn Lane Width must be co-populated. This did not occur for .58% of left turns and 1.12% of right turns. If a Road Characteristic segment has a turn type, then it must have a turn width. If it has a turn width, then it must have a turn type.
    - In almost 42% of Road Characteristics, the Addition Date has been populated, but the Addition Document Type is not populated. In many cases, it appears that the Addition Date has been populated with a legal (after 12/31/1930), but illogical date. These dates should be revisited.
    - In almost 42% of Road Characteristics, the Addition Date has been populated, but the Addition Document ID is not populated. In many cases, it appears that the Addition Date has been populated with a legal (after 12/31/1930), but illogical date. These dates should be revisited.
    - In more than 27% of Road Characteristics, the Addition Date has occurred before the original Improvement Date if the Improvement Date has been populated. Once again, it looks like the Addition Date has been populated with a legal (after 12/31/1930), but illogical date.
    - The addition date (ADTN_DT) contains values on 12/31 in years after 1930, which are valid, but unreasonable values. These should be cleaned up so the prior 3 errors do not occur in the future.

8. Discussion

While only a sampling of features were checked for horizontal and attribute accuracy because of time and resource constraints, the NCCU Research Team is confident with the results and satisfied with the procedures used to compile them. With a project of this magnitude, it is important to recognize the potential for error in the assessment of these metrics which may ultimately affect the final score. Many of these factors were unknown to the NCCU Research Team and NC DOT before this project began. If a project of this type were to be performed again, these are some variables that should be addressed.

8.1 Potential for Error with Horizontal Accuracy Metric

Horizontal Accuracy was measured using a random sampling of **LRS_Arcs** based on a total number of eligible **LRS_Arcs** within each county. Sample sizes ranged from 50 in smaller counties such as Gates and Clay Counties up to 315 **LRS_Arcs** for Wake County.

- Depending upon the county, about 1.3 – 10.7% of all **LRS_Arcs** within the county were evaluated for horizontal accuracy. While the number of **LRS_Arcs** selected adhered to acceptable QA/QC standards, Horizontal Accuracy for each county may vary slightly depending upon the number and percentage of **LRS_Arcs** selected for each county.

- **LRS_Arcs** of any length were eligible to be evaluated for horizontal accuracy. In **LRS_Arcs** with very short lengths, there was very little variation of distance between the 3 points along the **LRS_Arc** and the 3 points digitized by NCCU staff along the corresponding correct road centerline. It is suggested that a minimum threshold length (.1 miles) be set for **LRS_Arcs** to be evaluated for horizontal accuracy.
- No consideration was given to traffic counts or the URBN_ID_CD attribute field to select from **LRS_Arcs** which may have more traffic, additions or attention from the NC DOT. Some counties such as Gates County had no randomly selected urban **LRS_Arcs** while 90.5% of all randomly selected **LRS_Arcs** in Mecklenburg County were classified as urban. In the future, the NC DOT may want to dictate a certain percentage or distribution of urban and rural **LRS_Arcs** be selected for assessment and evaluation.

- In a few cases when digitizing along the road centerline to determine the actual road centerline versus a point along the randomly selected **LRS_Arc**, the road centerline was obscured by trees or other obstacles (see diagram to right). A best guess was made to determine the road centerline.

8.2 Potential for Error with Attribute Accuracy Metric

Attribute Accuracy was measured by checking 23 attributes (highlighted in Table 2 in Appendix B) from a sampling of **LRS_Arcs** based on a total number of eligible **LRS_Arcs** within each county. Sample sizes ranged from 50 in smaller counties such as Gates and Clay Counties up to 315 **LRS_Arcs** for Wake County.

- Some of the randomly selected **LRS_Arcs** for an entire county did not contain right turn lanes, left turn lanes and medians. As a result the turn/median types and widths had scores of 100% or only a minimal number of features below acceptable QA/QC standards. It is suggested that future samples contain a minimum number (10, for example) of **LRS_Arcs** that contain right turn lanes, left turn lanes and medians.
- There are more than 19,000 structures located through the state. Some counties had a minimal number of randomly selected **LRS_Arcs** containing structures. It is suggested that future samples contain a minimum number (10, for example) of **LRS_Arcs** that contain structures based on the GIS data layer provided by the NC DOT.
- The speed limit attribute (SPD_LMT_TYP_CD) that was checked was based on a layer used to create the original **LRS_Arc** feature. Road ordnance data provided through TEAAS was difficult to geocode and spatially match with the randomly selected **LRS_Arcs**.

8.3 Potential for Error with Attribute Completeness Metric

Using custom Python code, attribute completeness was checked on all Road Characteristics to ensure they adhered to ARID documentation. Tyrrell County has 465

Road Characteristics while Wake County has more than 25,000 separate Road Characteristic segments.

- Attribute completeness was checked based on current ARID documentation. Data imported from the universe may contain legacy values that are no longer valid. They will have been marked incorrect.

8.4 Potential for Error with Final Score

Values for the individual scores between 0 and 100 for the 3 metrics (Horizontal Accuracy, Attribute Accuracy and Attribute Completeness) were linearly scaled based on the minimum and maximum scores for each metric.

- The score for Horizontal Accuracy was skewed based on outliers. Clay County has a Horizontal Accuracy of 18.82 feet with an average for all counties being 5.35 feet. As a result, only 8 counties earned a value below 50, corresponding to the range midpoint of 9.71 feet. If Clay County's Horizontal Accuracy were lower, there would be more variation in the Horizontal Accuracy score and more values below 50.
- Each of the 3 metrics were equally weighted (1/3 for each metric) to compute a final score. If these metrics were weighted differently, the final scores (Map 10) would change accordingly.

9. Implementation Plan

The impetus for this project was to give the NC DOT a quantitative assessment of data quality utilized by the public on an everyday basis. The end goal was to turn information gleaned from GIS data into knowledge from which decisions can be made and action can be taken. As QA/QC technologies have vastly improved since the original NC DOT GIS database was created, both a proactive and retroactive approach has been taken by the NC DOT to ensure data quality now and in the future. The NC DOT is working on a photo-revision project to ensure that legacy GIS data better align with high-resolution imagery while eliminating old or outdated data, thus improving horizontal accuracy. Approximately fifteen of North Carolina's 100 counties have been completed.

To improve attribute accuracy and completeness, the NC DOT has custom software to ensure that proper domains and required fields are populated with the correct values moving forward. However, this software does not correct mistakes or oversights previously made. As part of the attribute completion assessment concluded by the research team, all 1,123,350 attribute completion errors were written to an error log, identifying each error by county name, G1FtSeg attribute (a unique identifier), error number and a short description of the error. A copy of this log was given to the NC DOT as they plan to plan to address these errors in the future. Lastly, a summary of attribute errors by county was provided to the NC DOT. The NC DOT plans to prioritize the most incomplete and incorrect attributes such as the RW_WID (Right of Way) and STRCTR_CD (Structure Type) on a county by county basis to complement attribute accuracy errors as a result of incomplete field attribution.

10. Acknowledgements

Dr. Timothy Mulrooney and the NCCU Research Team wishes to thank the NC DOT for the opportunity to work on a project of this magnitude. This project has provided valuable work experience for students getting ready to enter the workforce. NCCU is also grateful for input provided throughout the course of this project. While not fully complete, those who helped

facilitate this project included: Jonathan Arnold, Ashley Clowes, Tim Sheldon, Faith Johnson and Mustansir Kadibhai. Timothy Mulrooney and NCCU are happy to follow-up on this work or assist those who have questions about this document. His contact information is below:

Timothy Mulrooney
Department of Environmental, Earth and Geospatial Sciences
North Carolina Central University
1246 Mary M. Townes Science Complex
1801 Fayetteville St., Durham, NC, 27707
Phone: 919-530-6575
E-mail: tmulroon@nccu.edu

11. References Cited

Aronoff, Stan. (1989). *Geographic Information Systems: A Management Perspective*. Ottawa: WDL Publications.

Bierkens, M. F. P. & Burrough, P.A. (1993). The indicator approach to categorical soil data. *Journal of Soil Science, 44*, 361-368.

Bedard, Y. (1987). Uncertainties in land information systems databases. *Proceedings of the 8th International Symposium on Computer Assisted Cartography (Auto Carto 8)*, (pp. 175-184). Baltimore, MD.

Broome, F.R., & Meixler, D.B. (1990). The TIGER database structure. *Cartography and Geographic Information Science, 17(1)*, 39-47.

Buttenfield, B.P. (1993). Representing data quality. *Cartographica, 30(2-3)*, 1-7.

Caspary, W. and R. Scheuring. (1993). Positional accuracy in spatial databases. *Computers, Environment and Urban Systems 17*, 103-110.

Chapman, A.D. (2005). *Principles of Data Quality, version 1.0*. Report for the Global Biodiversity Information Facility, Copenhagen.

Devillers, R., Bédard, Y. & Jeansoulin, R. (2005). Multidimensional management of geospatial data quality information for its dynamic use within Geographical Information Systems. *Photogrammetric Engineering and Remote Sensing, 71(2)*, 205–215.

Devillers R, Y., Bédard, R. Jeansoulin, & Moulin, B. (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographic Information Sciences, 21(3)*, 261 – 282.

Epstein, E. F. (1988). Litigation over information: The use and misuse of maps. *Proceedings, IGIS: The Research Agenda* 1 (pp. 177-184). Washington, D.C.: NASA.

FGDC (Federal Geographic Data Committee). (1998). *National Standard for Spatial Data Accuracy*. Washington D.C.: Federal Geographic Data Committee.

Foody, G. M. (2005). Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing, 26*, 1217-1228.

Howard, D. and MacEachren, A.M. (1996). Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Systems, 23*, 59–77.

MacDonald, H. (2006). The American Community Survey: Warmer (more current), but fuzzier (less precise) than the decennial census. *Journal of the American Planning Association. American Planning Association, 72(4)*, 491-503.

MacEachren, A.M. (1992). Visualizing uncertain information. *Cartographic Perspective, 13*, 10-19.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M. & Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science, 32(3)*, 139-160.

Openshaw, S. (1989). Learning to live with errors in spatial databases. In: Goodchild, M. F. and S. Gopal (Eds.), *Accuracy of Spatial Databases* (pp. 263-276). London: Taylor and Francis.

Sinton, D. (1978). The inherent structure of information as a constraint to analysis. In: D. Dutton (ed.), *Harvard papers on geographic information systems*. Reading, U.K.: Addison Wesley.

Smith, J. H., Wickham, J. D., Stehman, S. V. & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment, 84*, 342-349.

Taulbee, S.M. (1996). Implementing data quality systems in biomedical records. In S.C. Gad, and S.M. Taulbee (Eds.) *Handbook of data recording, maintenance, and management for the biomedical sciences* (pp. 47-75). Boca Raton: CRC Press.

U.S. Census Bureau. (2008). *A compass for understanding and using American Community Survey data: What general data users need to know*. U.S. Government Printing Office, Washington, DC.

USGS. (1947). *National mapping accuracy standards*. Reston, Virginia: United States Geologic Survey

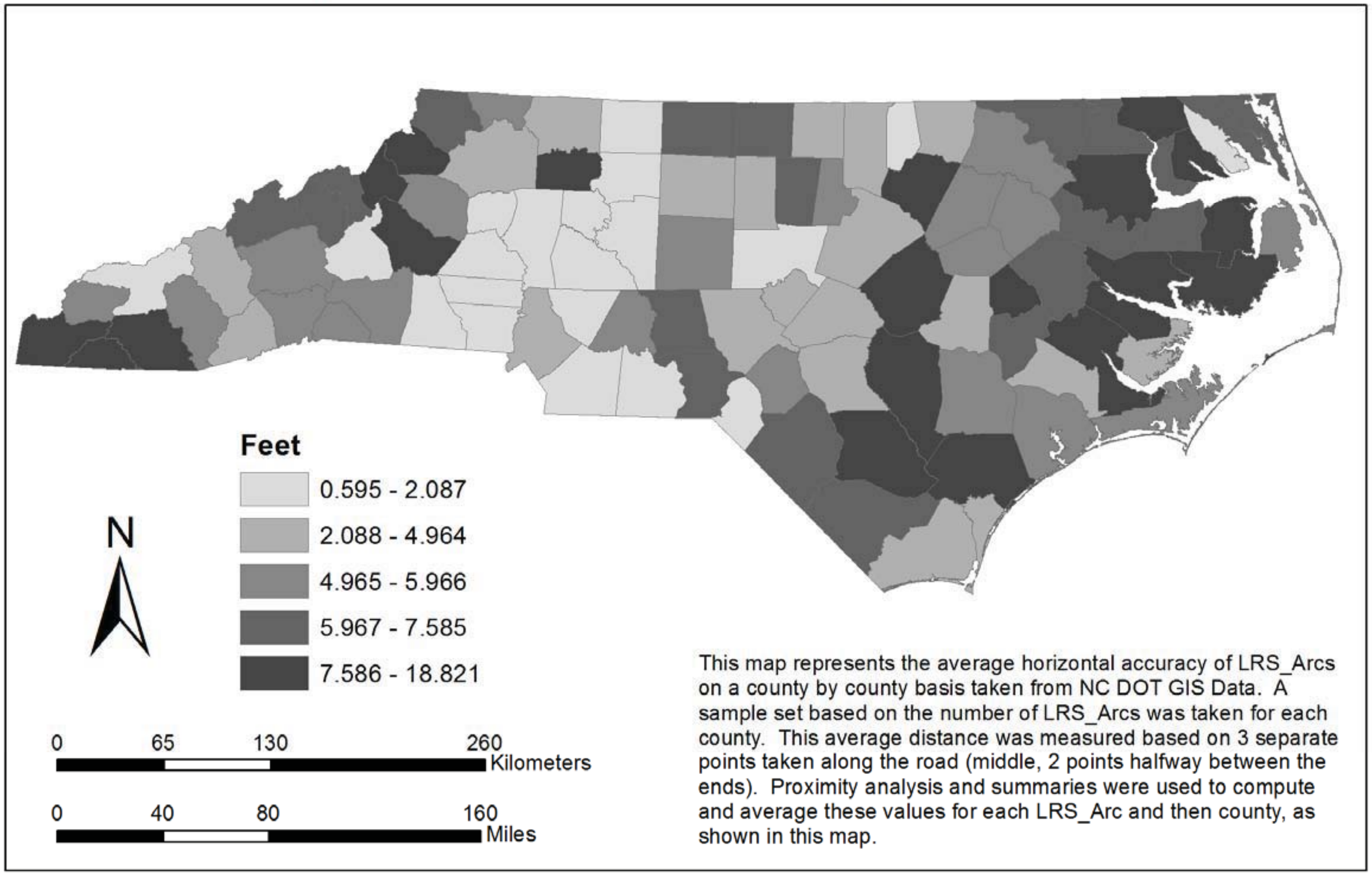
Wong, D.W. & Sun, M. (2013). Handling data quality information of survey data in GIS: A case of using the American Community Survey data. *Spatial Demography*, *1(1)*: 3-16.

Xiao, N., C., Calder, A. & Armstrong, M. P. (2007). Assessing the effect of attribute uncertainty on the robustness of choropleth map classification. *International Journal of Geographical Information Science, 21(2)*, 121-144.

Zandbergen, P.A. (2008). Positional accuracy of spatial data: Non-normal distributions and a critique of the National Standard for Spatial Data Accuracy. *Transactions in GIS, 12(1)*, 103-130.

Appendix A: Maps

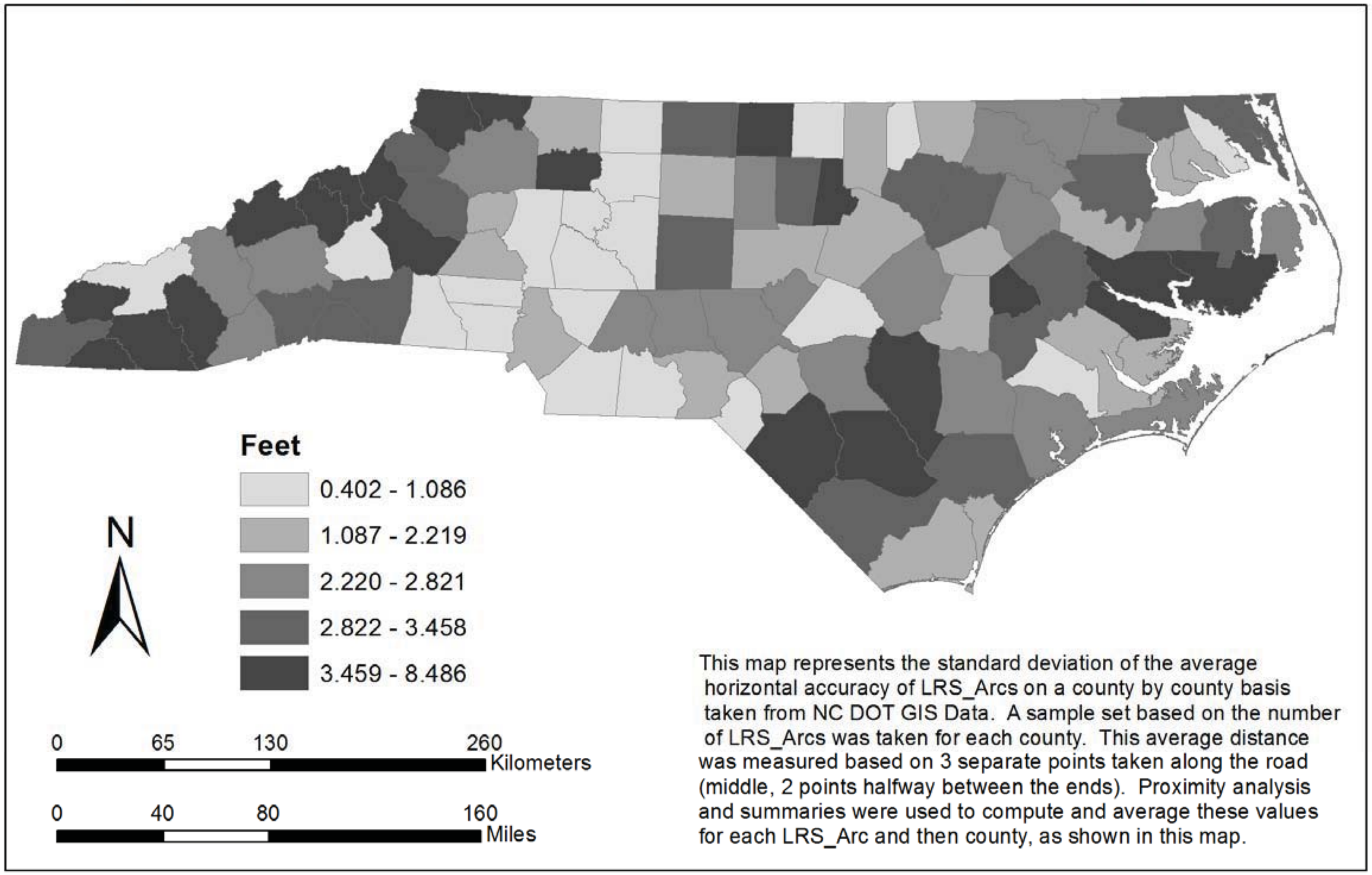
Map 1: Average Horizontal Accuracy by County



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

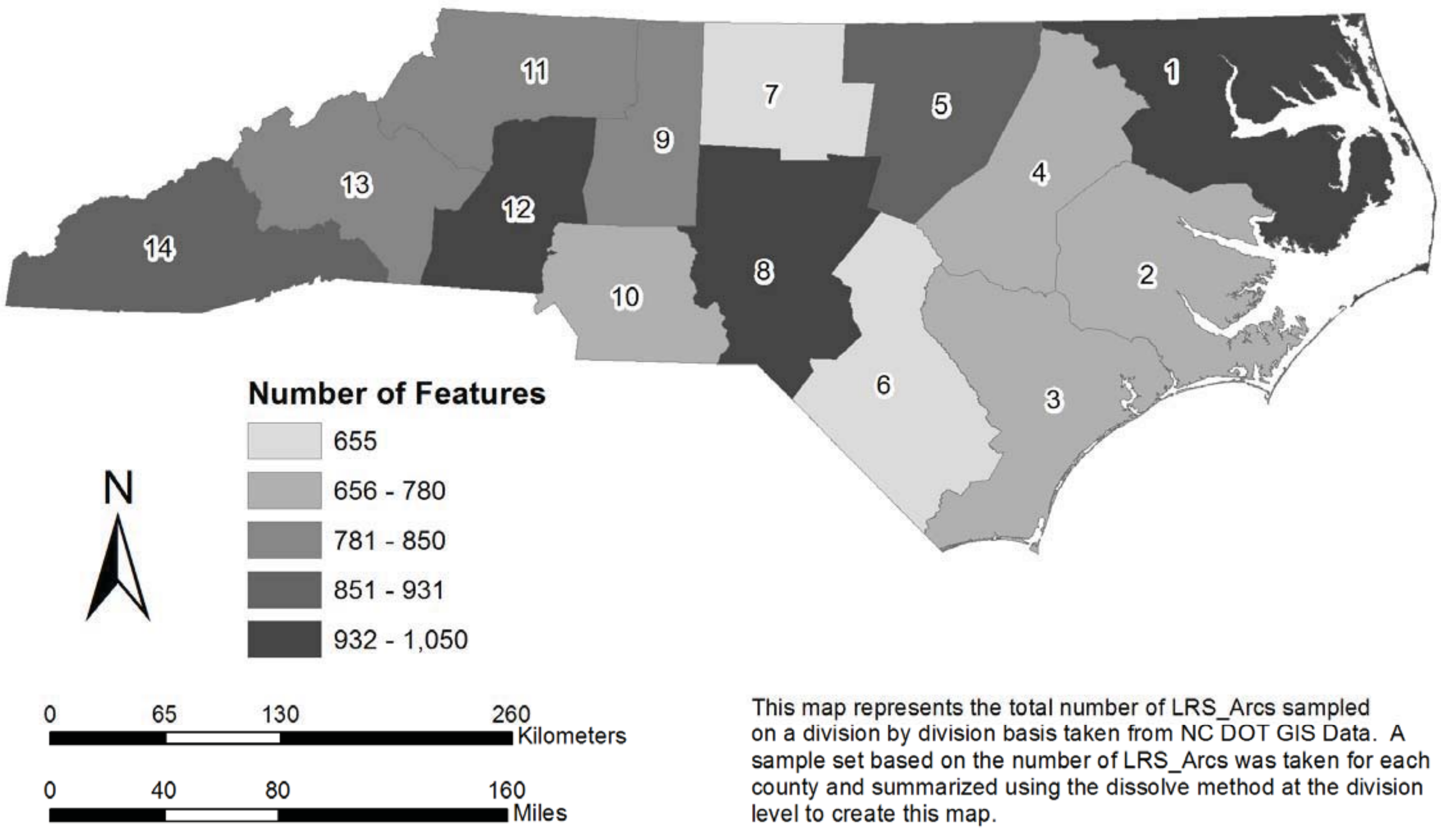
Map 2: Standard Deviation of Average Horizontal Accuracy by County



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 3: Number of Features by Division

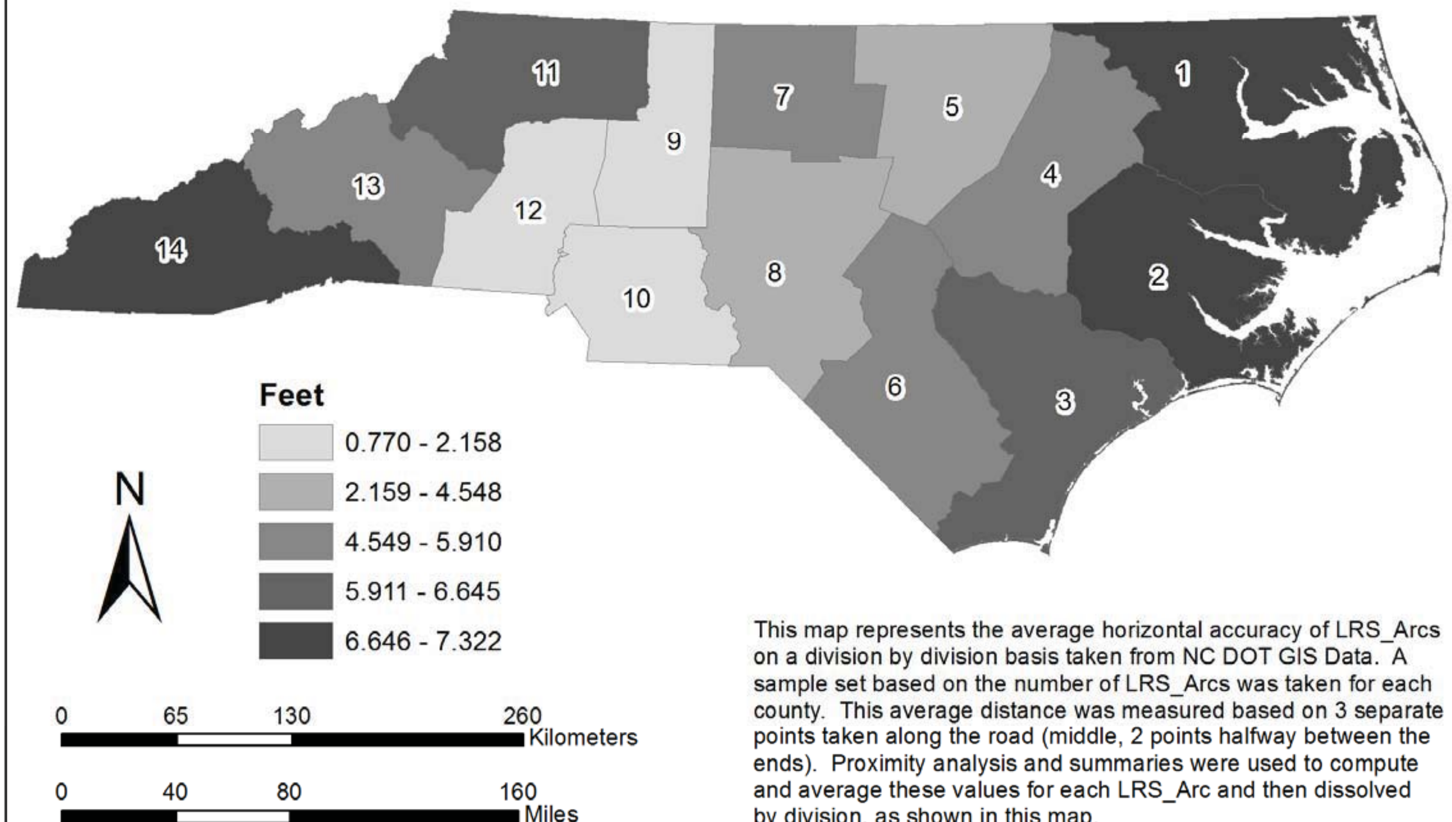


This map represents the total number of LRS_Arcs sampled on a division by division basis taken from NC DOT GIS Data. A sample set based on the number of LRS_Arcs was taken for each county and summarized using the dissolve method at the division level to create this map.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 4: Average Horizontal Accuracy by Division

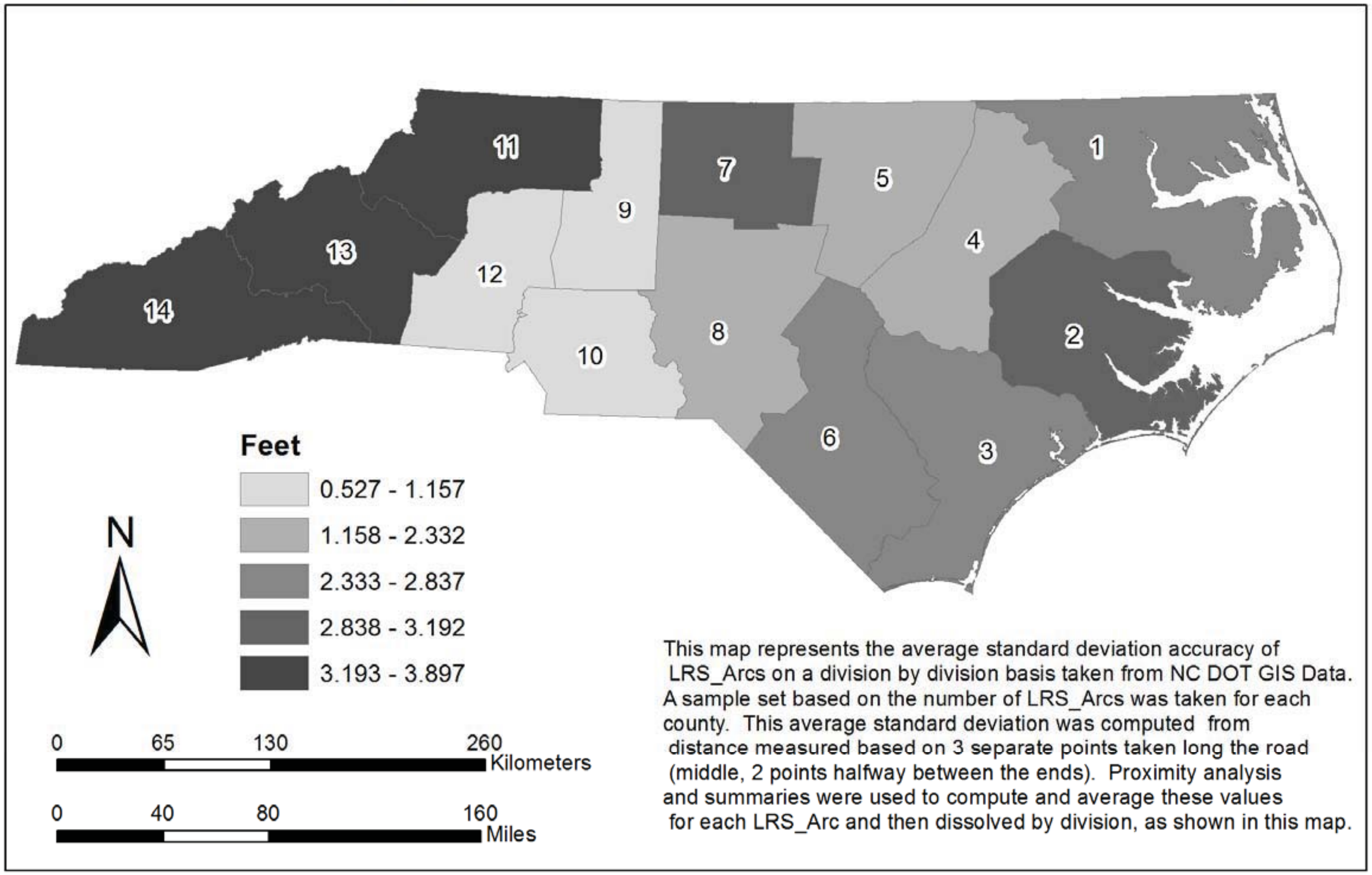


This map represents the average horizontal accuracy of LRS_Arcs on a division by division basis taken from NC DOT GIS Data. A sample set based on the number of LRS_Arcs was taken for each county. This average distance was measured based on 3 separate points taken along the road (middle, 2 points halfway between the ends). Proximity analysis and summaries were used to compute and average these values for each LRS_Arc and then dissolved by division, as shown in this map.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

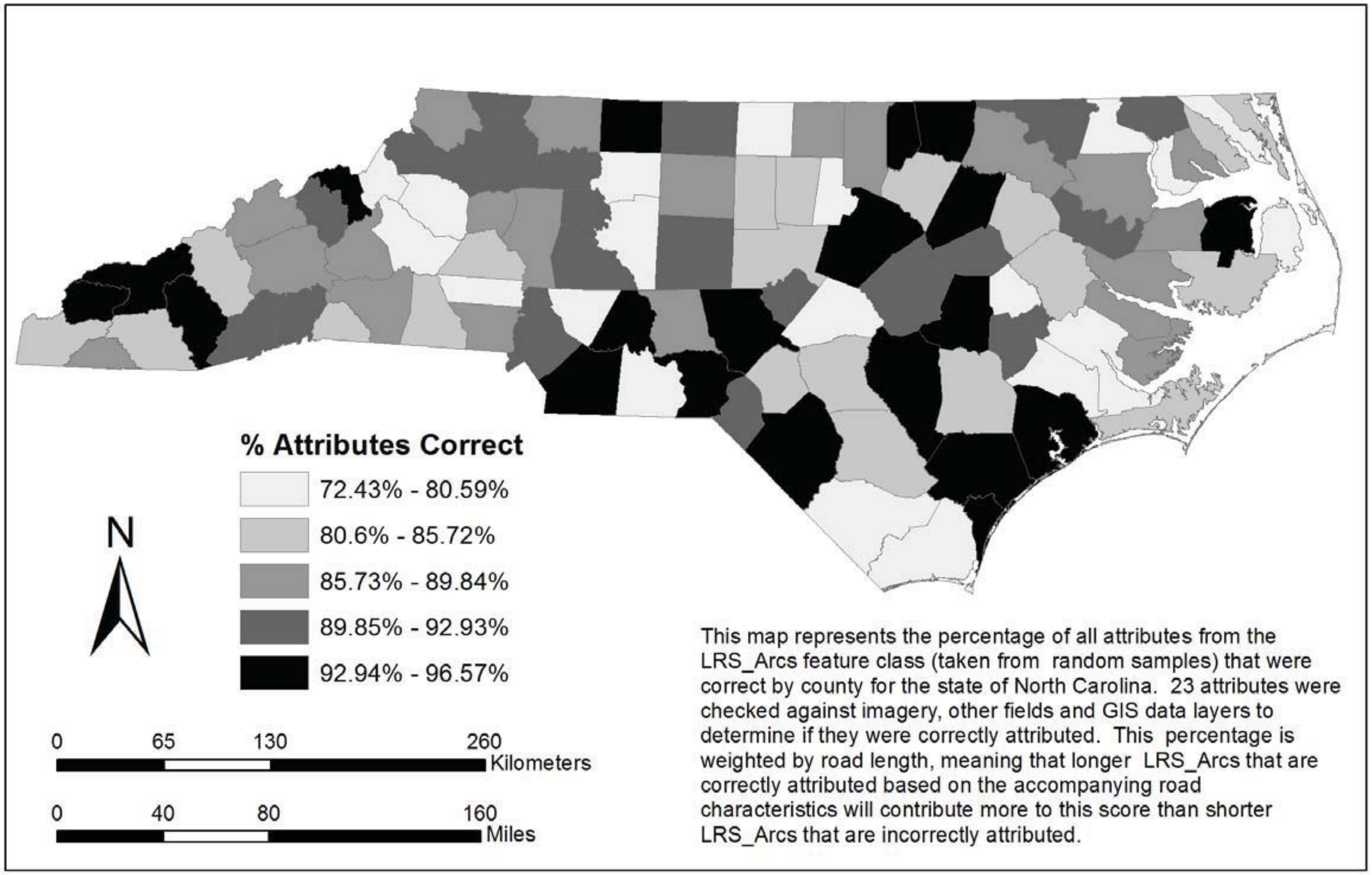
Map 5: Standard Deviation by Division



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

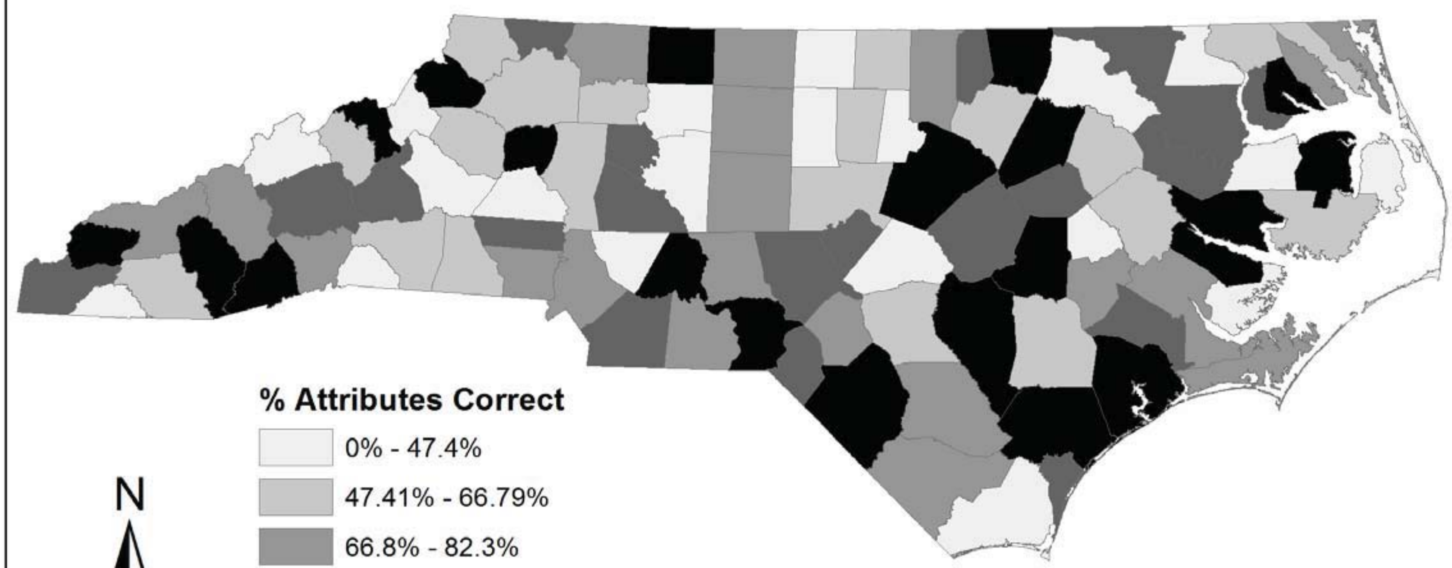
Map 6: Total Attribute Accuracy by County



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 6a: Median Type Attribute Accuracy by County

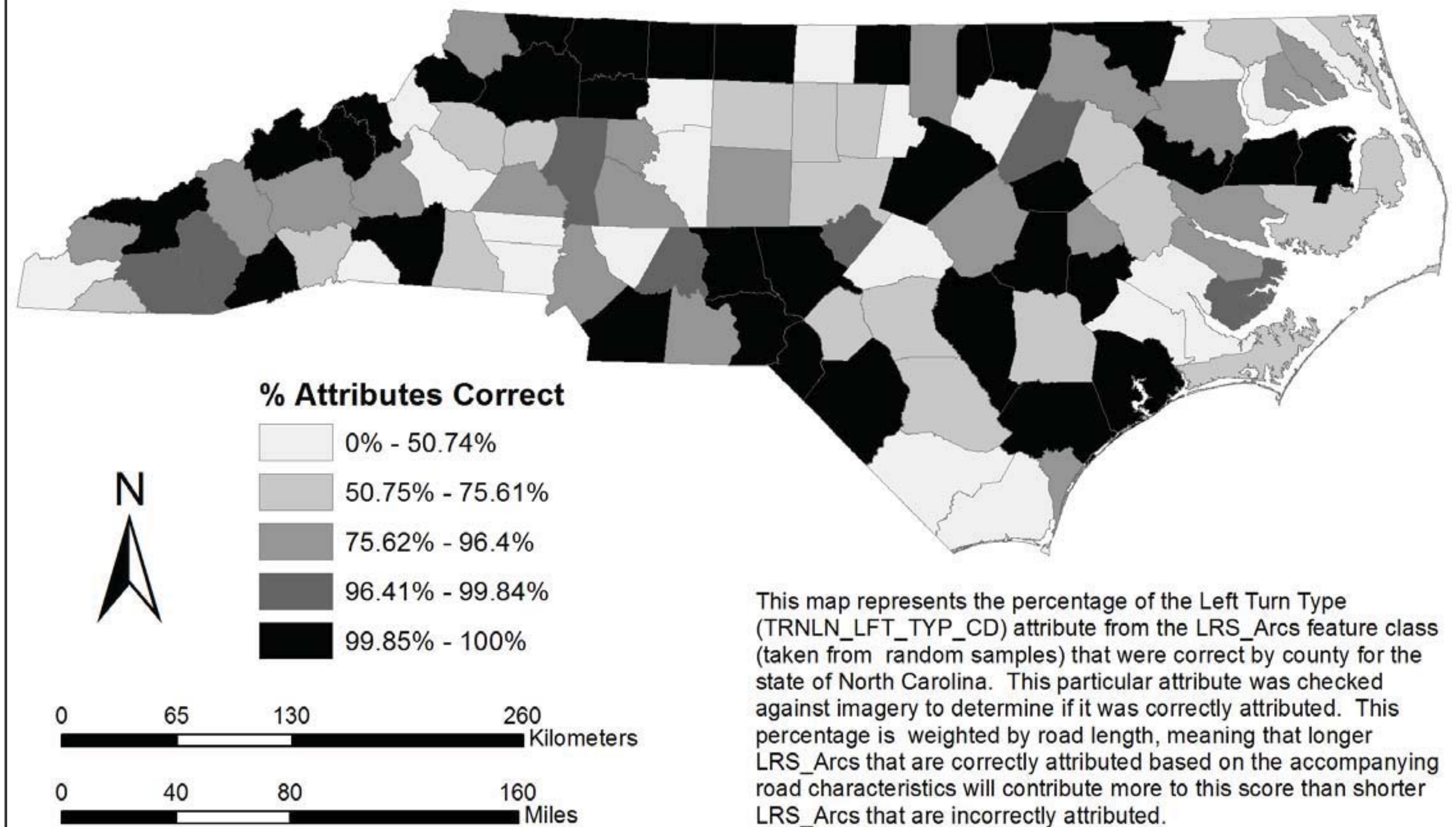


This map represents the percentage of the Median Type (MDN_TYP_CD) attribute from the LRS_Arcs feature class (taken from random samples) that were correct by county for the state of North Carolina. This particular attribute was checked against imagery to determine if it was correctly attributed. This percentage is weighted by road length, meaning that longer LRS_Arcs that are correctly attributed based on the accompanying road characteristics will contribute more to this score than shorter LRS_Arcs that are incorrectly attributed.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 6b: Left Turn Type Attribute Accuracy by County

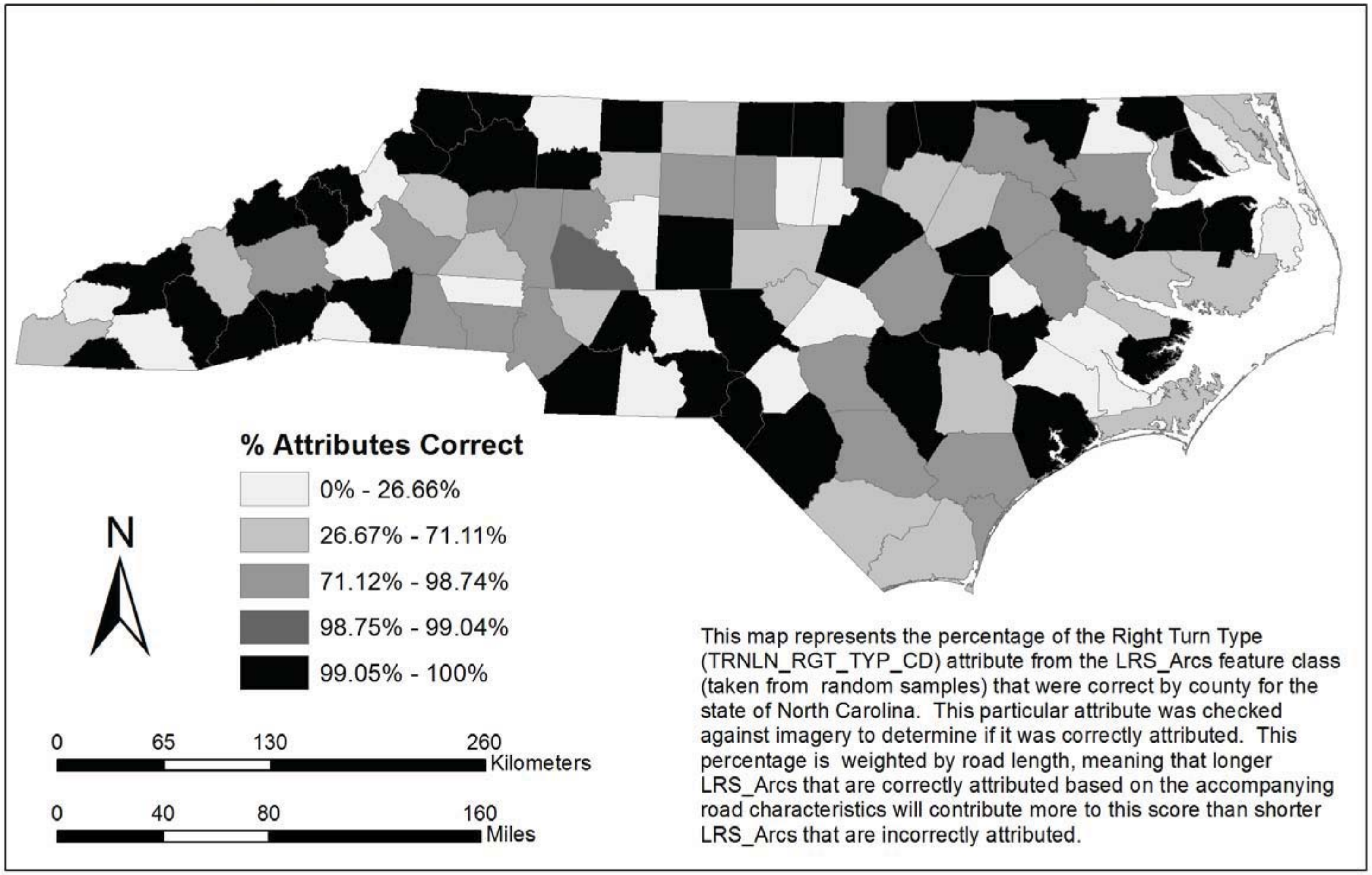


This map represents the percentage of the Left Turn Type (TRNLN_LFT_TYP_CD) attribute from the LRS_Arcs feature class (taken from random samples) that were correct by county for the state of North Carolina. This particular attribute was checked against imagery to determine if it was correctly attributed. This percentage is weighted by road length, meaning that longer LRS_Arcs that are correctly attributed based on the accompanying road characteristics will contribute more to this score than shorter LRS_Arcs that are incorrectly attributed.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

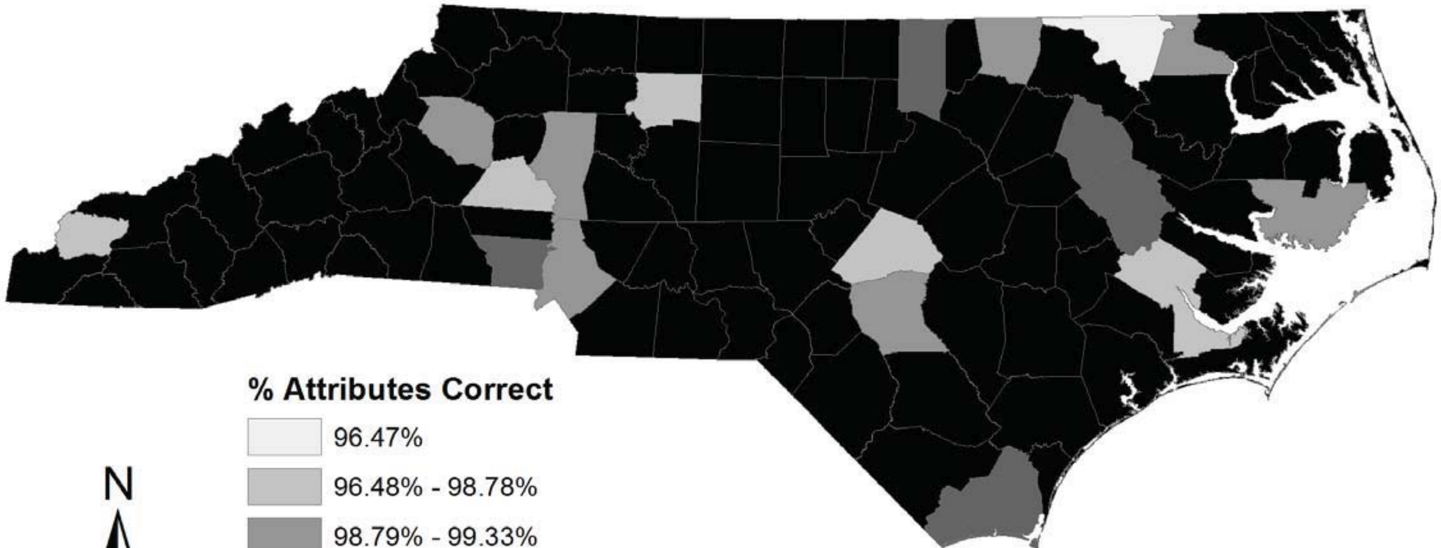
Map 6c: Right Turn Type Attribute Accuracy by County



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 6d: Number of Lanes Attribute Accuracy by County

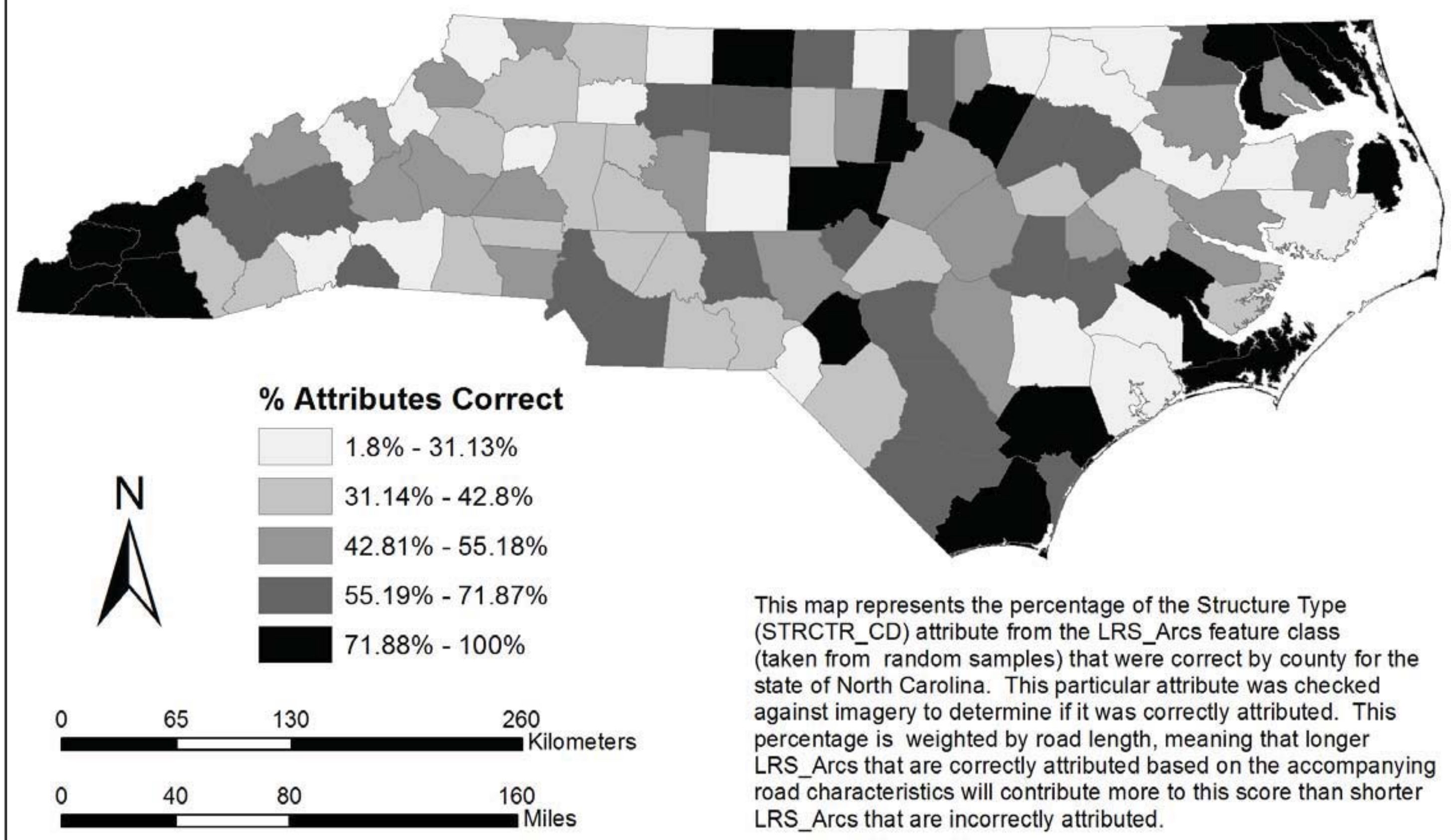


This map represents the percentage of the Number of Lanes (NBR_LANE_QTY) attribute from the LRS_Arcs feature class (taken from random samples) that were correct by county for the state of North Carolina. This particular attribute was checked against imagery to determine if it was correctly attributed. This percentage is weighted by road length, meaning that longer LRS_Arcs that are correctly attributed based on the accompanying road characteristics will contribute more to this score than shorter LRS_Arcs that are incorrectly attributed.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 6e: Structure Type Attribute Accuracy by County

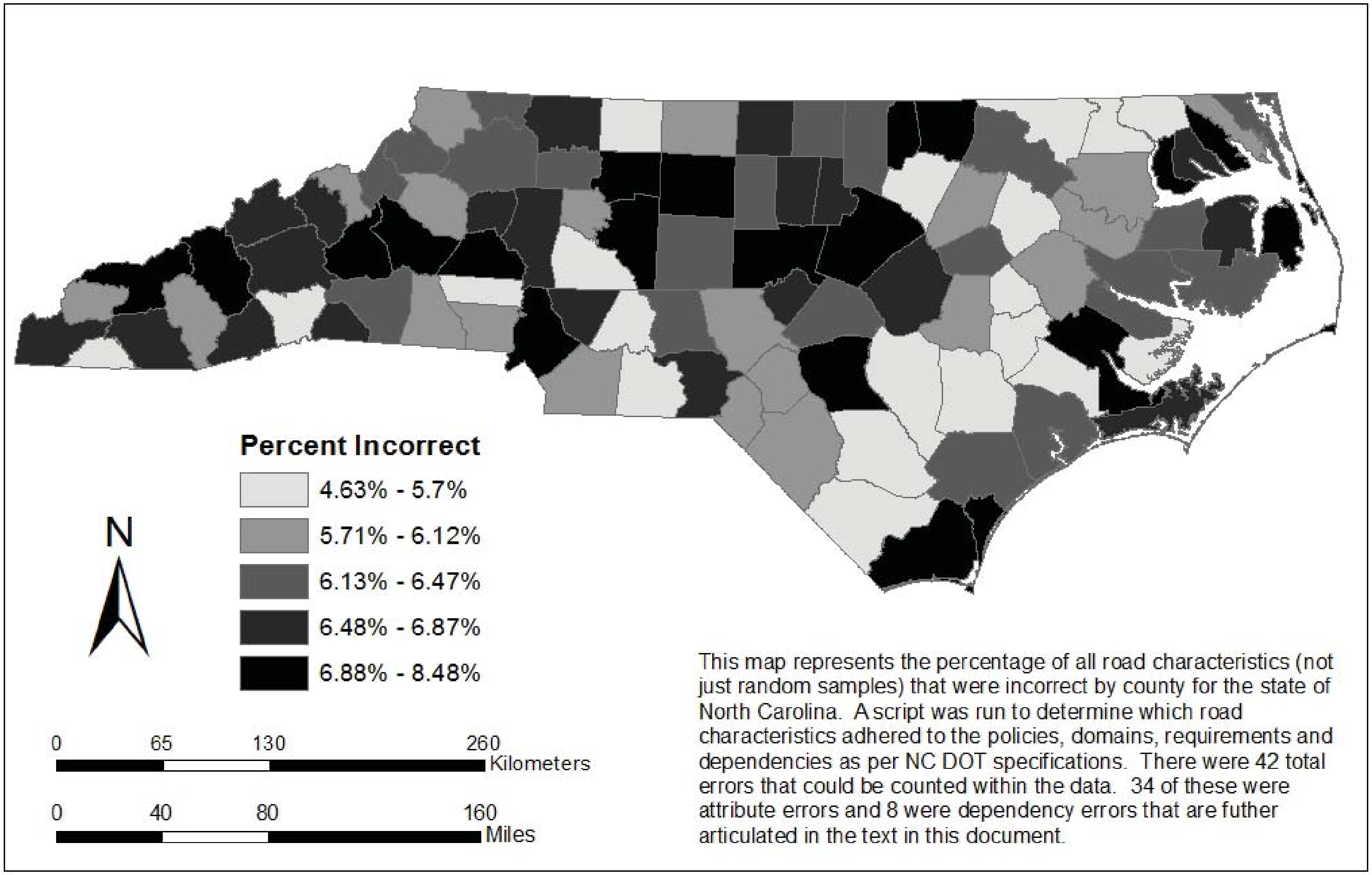


This map represents the percentage of the Structure Type (STRCTR_CD) attribute from the LRS_Arcs feature class (taken from random samples) that were correct by county for the state of North Carolina. This particular attribute was checked against imagery to determine if it was correctly attributed. This percentage is weighted by road length, meaning that longer LRS_Arcs that are correctly attributed based on the accompanying road characteristics will contribute more to this score than shorter LRS_Arcs that are incorrectly attributed.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

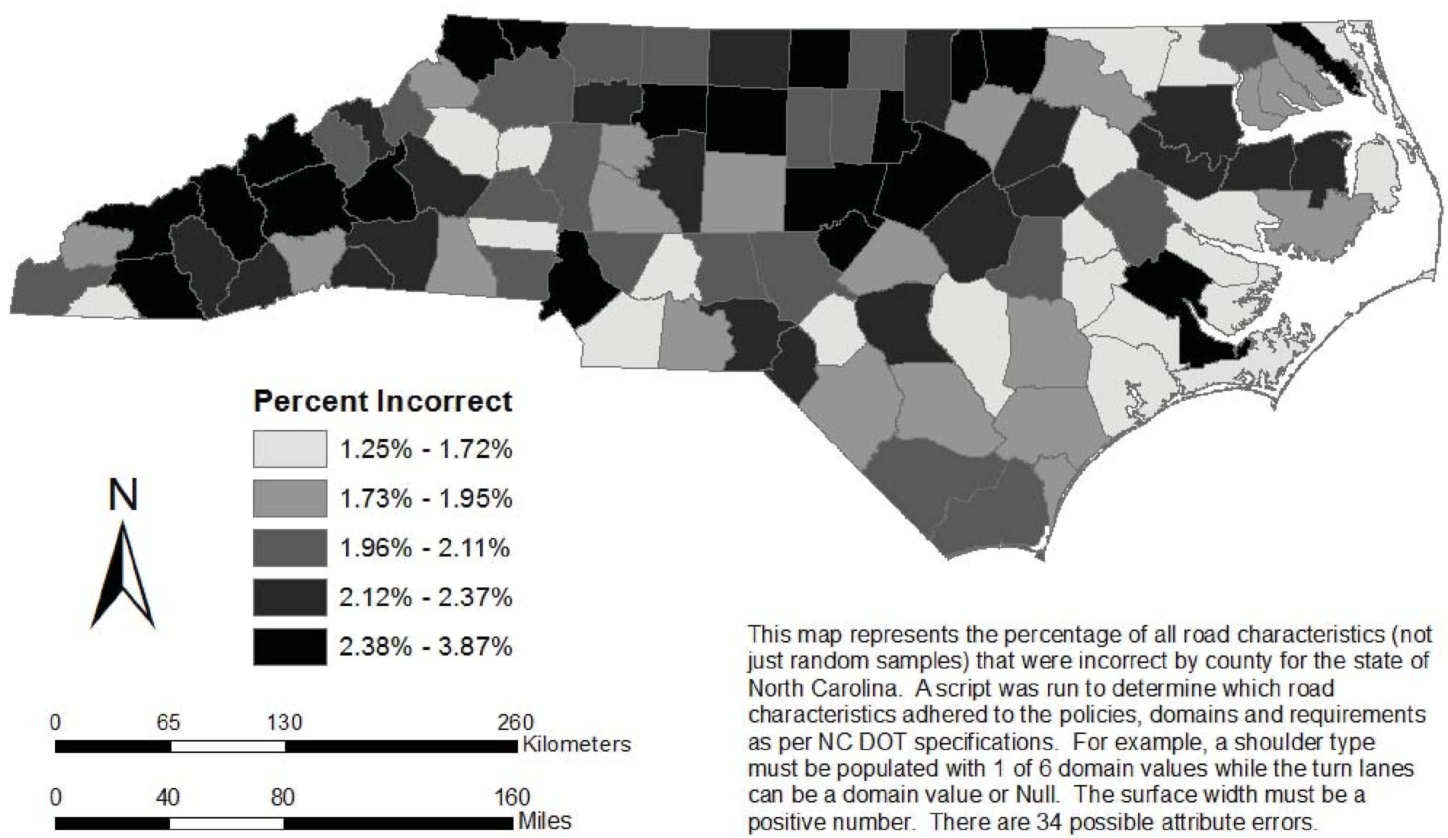
Map 7: Total Attribute and Dependency Completeness by County



Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 8: Total Attribute Completeness by County

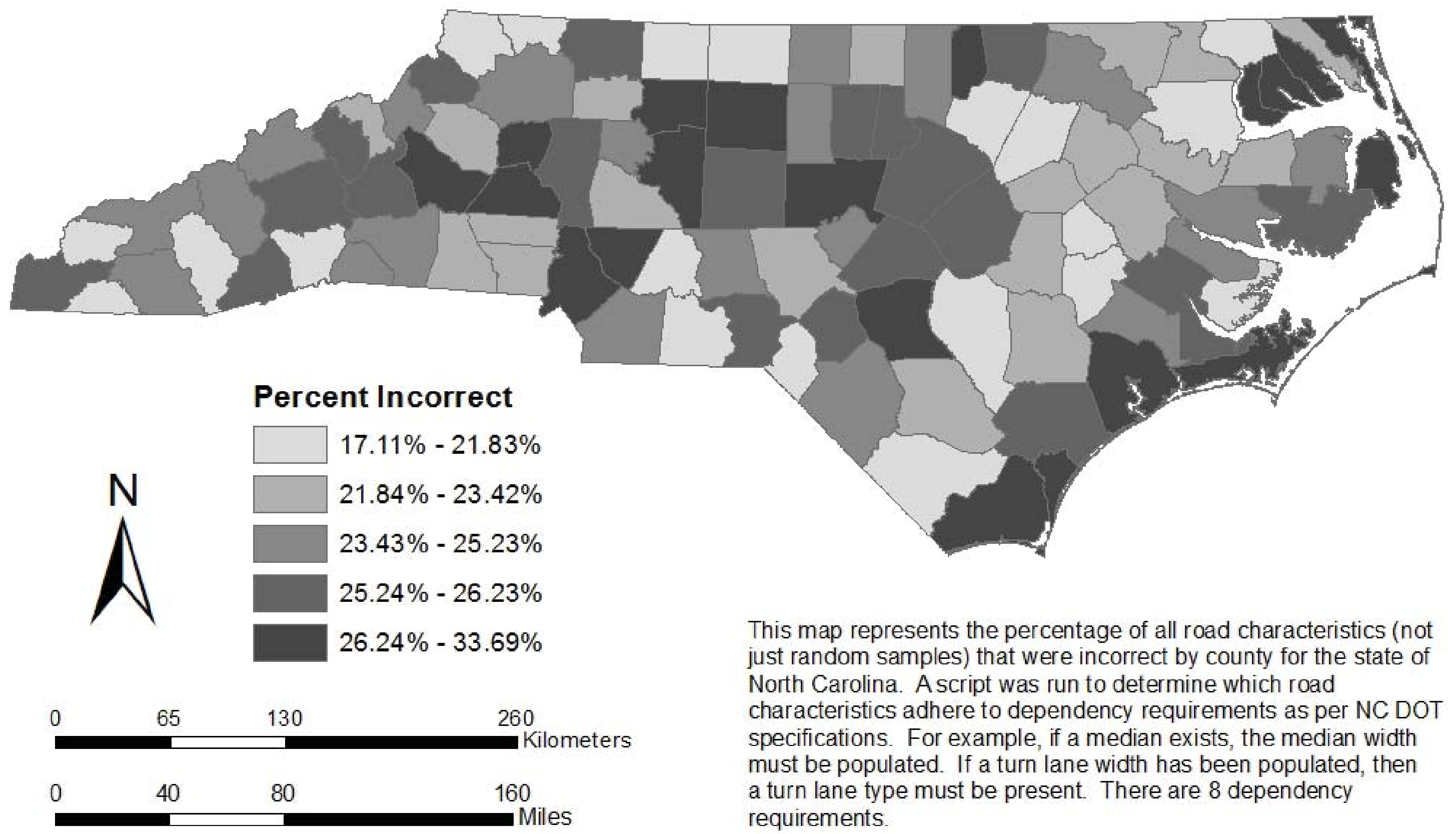


This map represents the percentage of all road characteristics (not just random samples) that were incorrect by county for the state of North Carolina. A script was run to determine which road characteristics adhered to the policies, domains and requirements as per NC DOT specifications. For example, a shoulder type must be populated with 1 of 6 domain values while the turn lanes can be a domain value or Null. The surface width must be a positive number. There are 34 possible attribute errors.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data*.

Map 9: Total Dependency Completeness by County

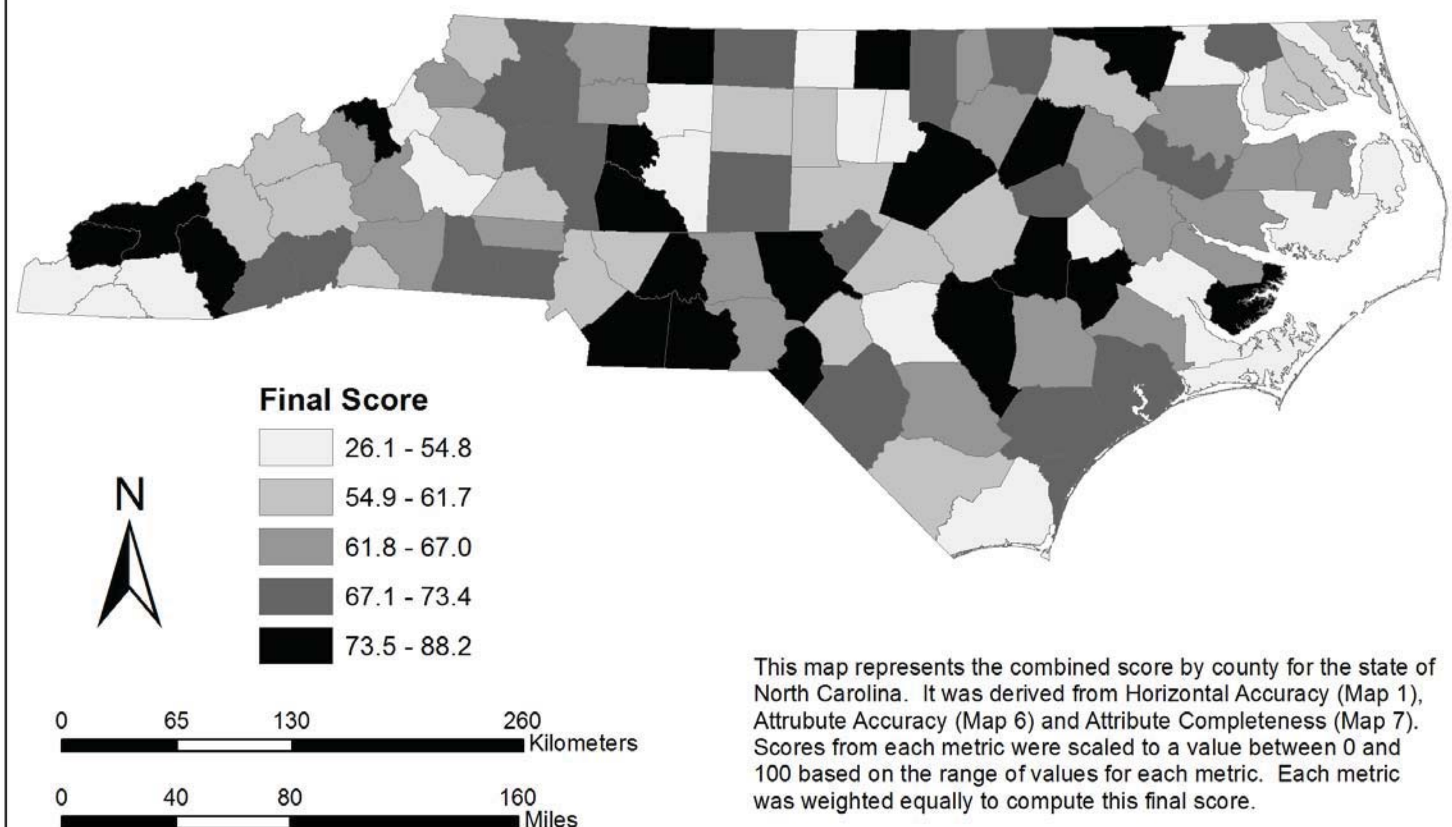


This map represents the percentage of all road characteristics (not just random samples) that were incorrect by county for the state of North Carolina. A script was run to determine which road characteristics adhere to dependency requirements as per NC DOT specifications. For example, if a median exists, the median width must be populated. If a turn lane width has been populated, then a turn lane type must be present. There are 8 dependency requirements.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Map 10: Final Score by County



This map represents the combined score by county for the state of North Carolina. It was derived from Horizontal Accuracy (Map 1), Attrubute Accuracy (Map 6) and Attribute Completeness (Map 7). Scores from each metric were scaled to a value between 0 and 100 based on the range of values for each metric. Each metric was weighted equally to compute this final score.

Timothy Mulrooney
Principal Investigator
tmulroon@nccu.edu
(919) 530 - 6575

This map was created by the Department of Environmental, Earth and Geospatial Sciences at North Carolina Central University. Funding for this project was provided by the North Carolina Department of Transportation Research Grant Program (#2014-02) under the title *Applying QA/QC Procedures to Quantitatively Measure the Quality of NCDOT GIS Data.*

Appendix B: Tables

Table 1: Summary of Randomly Selected LRS_Arcs Used in Sampling Process

County	< .01 miles	.01 - .1 miles	.1 - 1 mile	1 - 1.5 miles	1.5 - 2 miles	2+ miles	Total	Avg. Length (Ft.)
Alamance	1 (0.8%)	23 (18.4%)	88 (70.4%)	9 (7.2%)	3 (2.4%)	1 (0.8%)	125 (100%)	2312.14
Alexander	4 (3.2%)	28 (22.4%)	86 (68.8%)	3 (2.4%)	3 (2.4%)	1 (0.8%)	125 (100%)	1795.61
Alleghany	2 (2.5%)	7 (8.75%)	54 (67.5%)	12 (15.0%)	1 (1.25%)	4 (5.0%)	80 (100%)	3646.13
Anson	2 (2.5%)	12 (15.0%)	46 (57.5%)	13 (16.25%)	6 (7.5%)	1 (1.25%)	80 (100%)	3345.85
Ashe	0 (0.0%)	10 (12.5%)	46 (57.5%)	13 (16.25%)	8 (10.0%)	3 (3.75%)	80 (100%)	3839.93
Avery	2 (2.5%)	7 (8.75%)	57 (71.25%)	7 (8.75%)	6 (7.5%)	1 (1.25%)	80 (100%)	3039.98
Beaufort	3 (2.4%)	23 (18.4%)	74 (59.2%)	10 (8.0%)	5 (4.0%)	10 (8.0%)	125 (100%)	3757.59
Bertie	5 (6.25%)	8 (10.0%)	43 (53.75%)	10 (12.5%)	7 (8.75%)	7 (8.75%)	80 (100%)	4142.48
Bladen	4 (5.0%)	5 (6.25%)	49 (61.25%)	10 (12.5%)	8 (10.0%)	4 (5.0%)	80 (100%)	3986.67
Brunswick	0 (0.0%)	31 (24.8%)	77 (61.6%)	8 (6.4%)	2 (1.6%)	7 (5.6%)	125 (100%)	2647.6
Buncombe	3 (1.5%)	64 (32.0%)	117 (58.5%)	8 (4.0%)	5 (2.5%)	3 (1.5%)	200 (100%)	1823.07
Burke	1 (0.8%)	30 (24.0%)	80 (64.0%)	8 (6.4%)	5 (4.0%)	1 (0.8%)	125 (100%)	2112.27
Cabarrus	3 (2.4%)	38 (30.4%)	78 (62.4%)	4 (3.2%)	2 (1.6%)	0 (0.0%)	125 (100%)	1664.61
Caldwell	0 (0.0%)	24 (19.2%)	86 (68.8%)	6 (4.8%)	5 (4.0%)	4 (3.2%)	125 (100%)	2696.55
Camden	3 (6.0%)	13 (26.0%)	27 (54.0%)	4 (8.0%)	1 (2.0%)	2 (4.0%)	50 (100%)	2267.53
Carteret	2 (2.5%)	23 (28.75%)	49 (61.25%)	3 (3.75%)	1 (1.25%)	2 (2.5%)	80 (100%)	2052.55
Caswell	1 (1.25%)	18 (22.5%)	46 (57.5%)	5 (6.25%)	7 (8.75%)	3 (3.75%)	80 (100%)	3256.37
Catawba	3 (1.5%)	63 (31.5%)	125 (62.5%)	6 (3.0%)	2 (1.0%)	1 (0.5%)	200 (100%)	1489.72
Chatham	2 (1.6%)	26 (20.8%)	73 (58.4%)	14 (11.2%)	7 (5.6%)	3 (2.4%)	125 (100%)	2793.31
Cherokee	0 (0.0%)	15 (18.75%)	54 (67.5%)	5 (6.25%)	4 (5.0%)	2 (2.5%)	80 (100%)	2753.99
Chowan	1 (1.25%)	26 (32.5%)	45 (56.25%)	5 (6.25%)	3 (3.75%)	0 (0.0%)	80 (100%)	2087.3
Clay	0 (0.0%)	7 (14.0%)	33 (66.0%)	6 (12.0%)	2 (4.0%)	2 (4.0%)	50 (100%)	3198.58
Cleveland	1 (0.5%)	46 (23.0%)	134 (67.0%)	12 (6.0%)	6 (3.0%)	1 (0.5%)	200 (100%)	2093.65
Columbus	0 (0.0%)	20 (16.0%)	74 (59.2%)	19 (15.2%)	6 (4.8%)	6 (4.8%)	125 (100%)	3732.86
Craven	2 (1.6%)	27 (21.6%)	81 (64.8%)	8 (6.4%)	5 (4.0%)	2 (1.6%)	125 (100%)	2505.51
Cumberland	6 (3.0%)	84 (42.0%)	95 (47.5%)	9 (4.5%)	5 (2.5%)	1 (0.5%)	200 (100%)	1426.31
Currituck	1 (1.25%)	18 (22.5%)	53 (66.25%)	4 (5.0%)	2 (2.5%)	2 (2.5%)	80 (100%)	2189.13
Dare	0 (0.0%)	24 (30.0%)	51 (63.75%)	3 (3.75%)	2 (2.5%)	0 (0.0%)	80 (100%)	1452.45
Davidson	3 (1.5%)	67 (33.5%)	122 (61.0%)	8 (4.0%)	0 (0.0%)	0 (0.0%)	200 (100%)	1376.63
Davie	3 (2.4%)	32 (25.6%)	76 (60.8%)	6 (4.8%)	6 (4.8%)	2 (1.6%)	125 (100%)	2128.47
Duplin	3 (2.4%)	9 (7.2%)	75 (60.0%)	23 (18.4%)	12 (9.6%)	3 (2.4%)	125 (100%)	4017.29

County	< .01 miles	.01 - .1 miles	.1 - 1 mile	1 - 1.5 miles	1.5 - 2 miles	2+ miles	Total	Avg. Length (Ft.)
Durham	6 (4.8%)	38 (30.4%)	77 (61.6%)	2 (1.6%)	2 (1.6%)	0 (0.0%)	125 (100%)	1334.35
Edgecombe	2 (2.5%)	12 (15.0%)	43 (53.75%)	14 (17.5%)	7 (8.75%)	2 (2.5%)	80 (100%)	3395.66
Forsyth	4 (2.0%)	72 (36.0%)	119 (59.5%)	3 (1.5%)	2 (1.0%)	0 (0.0%)	200 (100%)	1182.17
Franklin	6 (4.8%)	19 (15.2%)	74 (59.2%)	16 (12.8%)	5 (4.0%)	5 (4.0%)	125 (100%)	3037.55
Gaston	9 (4.5%)	70 (35.0%)	112 (56.0%)	4 (2.0%)	3 (1.5%)	2 (1.0%)	200 (100%)	1385.02
Gates	0 (0.0%)	8 (16.0%)	28 (56.0%)	5 (10.0%)	4 (8.0%)	5 (10.0%)	50 (100%)	4345.46
Graham	1 (2.0%)	4 (8.0%)	32 (64.0%)	7 (14.0%)	3 (6.0%)	3 (6.0%)	50 (100%)	4201.61
Granville	5 (4.0%)	17 (13.6%)	81 (64.8%)	11 (8.8%)	5 (4.0%)	6 (4.8%)	125 (100%)	3114.31
Greene	6 (7.5%)	10 (12.5%)	52 (65.0%)	10 (12.5%)	2 (2.5%)	0 (0.0%)	80 (100%)	2736.23
Guilford	2 (1.0%)	67 (33.5%)	120 (60.0%)	8 (4.0%)	1 (0.5%)	2 (1.0%)	200 (100%)	1613.11
Halifax	3 (2.4%)	38 (30.4%)	55 (44.0%)	13 (10.4%)	6 (4.8%)	10 (8.0%)	125 (100%)	3253.04
Harnett	2 (1.6%)	29 (23.2%)	76 (60.8%)	9 (7.2%)	5 (4.0%)	4 (3.2%)	125 (100%)	2559.61
Haywood	6 (4.8%)	29 (23.2%)	80 (64.0%)	7 (5.6%)	2 (1.6%)	1 (0.8%)	125 (100%)	1968.12
Henderson	2 (1.6%)	34 (27.2%)	80 (64.0%)	8 (6.4%)	0 (0.0%)	1 (0.8%)	125 (100%)	1877.05
Hertford	0 (0.0%)	8 (10.0%)	50 (62.5%)	10 (12.5%)	5 (6.25%)	7 (8.75%)	80 (100%)	3863.74
Hoke	2 (1.6%)	35 (28.0%)	80 (64.0%)	4 (3.2%)	3 (2.4%)	1 (0.8%)	125 (100%)	1861.04
Hyde	0 (0.0%)	11 (22.0%)	27 (54.0%)	4 (8.0%)	5 (10.0%)	3 (6.0%)	50 (100%)	3537.54
Iredell	1 (0.5%)	48 (24.0%)	132 (66.0%)	12 (6.0%)	7 (3.5%)	0 (0.0%)	200 (100%)	2062.67
Jackson	3 (2.4%)	21 (16.8%)	86 (68.8%)	4 (3.2%)	6 (4.8%)	5 (4.0%)	125 (100%)	2873.25
Johnston	1 (0.5%)	53 (26.5%)	127 (63.5%)	14 (7.0%)	4 (2.0%)	1 (0.5%)	200 (100%)	2153.57
Jones	2 (4.0%)	9 (18.0%)	23 (46.0%)	6 (12.0%)	2 (4.0%)	8 (16.0%)	50 (100%)	5581.15
Lee	5 (4.0%)	30 (24.0%)	82 (65.6%)	7 (5.6%)	1 (0.8%)	0 (0.0%)	125 (100%)	2003.15
Lenoir	1 (0.8%)	19 (15.2%)	85 (68.0%)	10 (8.0%)	8 (6.4%)	2 (1.6%)	125 (100%)	2823.84
Lincoln	2 (1.6%)	27 (21.6%)	91 (72.8%)	4 (3.2%)	1 (0.8%)	0 (0.0%)	125 (100%)	1667.76
Macon	2 (1.6%)	27 (21.6%)	84 (67.2%)	4 (3.2%)	3 (2.4%)	5 (4.0%)	125 (100%)	2642.7
Madison	0 (0.0%)	13 (16.25%)	53 (66.25%)	3 (3.75%)	4 (5.0%)	7 (8.75%)	80 (100%)	3797.98
Martin	1 (1.25%)	14 (17.5%)	47 (58.75%)	10 (12.5%)	2 (2.5%)	6 (7.5%)	80 (100%)	3481.56
McDowell	5 (4.0%)	38 (30.4%)	63 (50.4%)	11 (8.8%)	6 (4.8%)	2 (1.6%)	125 (100%)	2252.97
Mecklenburg	12 (6.0%)	74 (37.0%)	105 (52.5%)	4 (2.0%)	3 (1.5%)	2 (1.0%)	200 (100%)	1420.78
Mitchell	1 (1.25%)	11 (13.75%)	55 (68.75%)	5 (6.25%)	4 (5.0%)	4 (5.0%)	80 (100%)	3392.29
Montgomery	3 (3.75%)	16 (20.0%)	46 (57.5%)	9 (11.25%)	4 (5.0%)	2 (2.5%)	80 (100%)	2988.21
Moore	2 (1.6%)	23 (18.4%)	76 (60.8%)	16 (12.8%)	3 (2.4%)	5 (4.0%)	125 (100%)	2995.56
Nash	3 (2.4%)	28 (22.4%)	75 (60.0%)	12 (9.6%)	5 (4.0%)	2 (1.6%)	125 (100%)	2588.6
New Hanover	2 (1.6%)	61 (48.8%)	58 (46.4%)	4 (3.2%)	0 (0.0%)	0 (0.0%)	125 (100%)	906.5
Northampton	0 (0.0%)	10 (12.5%)	50 (62.5%)	9 (11.25%)	6 (7.5%)	5 (6.25%)	80 (100%)	3658.15

County	< .01 miles	.01 - .1 miles	.1 - 1 mile	1 - 1.5 miles	1.5 - 2 miles	2+ miles	Total	Avg. Length (Ft.)
Onslow	2 (1.6%)	42 (33.6%)	72 (57.6%)	3 (2.4%)	3 (2.4%)	3 (2.4%)	125 (100%)	1699.62
Orange	5 (4.0%)	29 (23.2%)	79 (63.2%)	9 (7.2%)	3 (2.4%)	0 (0.0%)	125 (100%)	2200.2
Pamlico	0 (0.0%)	6 (12.0%)	33 (66.0%)	5 (10.0%)	1 (2.0%)	5 (10.0%)	50 (100%)	3646.41
Pasquotank	2 (2.5%)	30 (37.5%)	43 (53.75%)	1 (1.25%)	2 (2.5%)	2 (2.5%)	80 (100%)	1740.55
Pender	2 (1.6%)	14 (11.2%)	83 (66.4%)	13 (10.4%)	9 (7.2%)	4 (3.2%)	125 (100%)	3444.16
Perquimans	0 (0.0%)	22 (27.5%)	46 (57.5%)	6 (7.5%)	3 (3.75%)	3 (3.75%)	80 (100%)	2569.28
Person	2 (2.5%)	17 (21.25%)	46 (57.5%)	7 (8.75%)	4 (5.0%)	4 (5.0%)	80 (100%)	3352.55
Pitt	1 (0.8%)	27 (21.6%)	78 (62.4%)	14 (11.2%)	2 (1.6%)	3 (2.4%)	125 (100%)	2545.83
Polk	4 (5.0%)	12 (15.0%)	47 (58.75%)	11 (13.75%)	5 (6.25%)	1 (1.25%)	80 (100%)	3057.33
Randolph	6 (3.0%)	55 (27.5%)	118 (59.0%)	8 (4.0%)	10 (5.0%)	3 (1.5%)	200 (100%)	1967.6
Richmond	2 (1.6%)	48 (38.4%)	63 (50.4%)	6 (4.8%)	2 (1.6%)	4 (3.2%)	125 (100%)	2199.64
Robeson	1 (0.8%)	18 (14.4%)	87 (69.6%)	8 (6.4%)	5 (4.0%)	6 (4.8%)	125 (100%)	2940.21
Rockingham	0 (0.0%)	24 (19.2%)	88 (70.4%)	8 (6.4%)	3 (2.4%)	2 (1.6%)	125 (100%)	2325.06
Rowan	5 (2.5%)	48 (24.0%)	127 (63.5%)	12 (6.0%)	8 (4.0%)	0 (0.0%)	200 (100%)	1971.97
Rutherford	1 (0.8%)	24 (19.2%)	81 (64.8%)	13 (10.4%)	4 (3.2%)	2 (1.6%)	125 (100%)	2600.62
Sampson	3 (2.4%)	22 (17.6%)	66 (52.8%)	18 (14.4%)	10 (8.0%)	6 (4.8%)	125 (100%)	3737.65
Scotland	4 (3.2%)	33 (26.4%)	76 (60.8%)	6 (4.8%)	5 (4.0%)	1 (0.8%)	125 (100%)	2218.67
Stanly	3 (2.4%)	18 (14.4%)	89 (71.2%)	8 (6.4%)	5 (4.0%)	2 (1.6%)	125 (100%)	2521.03
Stokes	0 (0.0%)	22 (17.6%)	83 (66.4%)	10 (8.0%)	6 (4.8%)	4 (3.2%)	125 (100%)	2864.25
Surry	1 (0.8%)	30 (24.0%)	81 (64.8%)	9 (7.2%)	2 (1.6%)	2 (1.6%)	125 (100%)	2199.45
Swain	1 (1.25%)	16 (20.0%)	52 (65.0%)	6 (7.5%)	4 (5.0%)	1 (1.25%)	80 (100%)	2463.72
Transylvania	3 (3.75%)	10 (12.5%)	60 (75.0%)	4 (5.0%)	0 (0.0%)	3 (3.75%)	80 (100%)	2400.92
Tyrrell	0 (0.0%)	5 (10.0%)	33 (66.0%)	2 (4.0%)	2 (4.0%)	8 (16.0%)	50 (100%)	5207.43
Union	1 (0.5%)	61 (30.5%)	121 (60.5%)	12 (6.0%)	3 (1.5%)	2 (1.0%)	200 (100%)	1835
Vance	0 (0.0%)	26 (32.5%)	46 (57.5%)	2 (2.5%)	4 (5.0%)	2 (2.5%)	80 (100%)	2234.74
Wake	7 (2.22%)	129 (40.95%)	166 (52.69%)	8 (2.53%)	3 (0.95%)	2 (0.63%)	315 (100%)	1344.4
Warren	2 (2.5%)	23 (28.75%)	35 (43.75%)	8 (10.0%)	7 (8.75%)	5 (6.25%)	80 (100%)	3170.16
Washington	1 (1.25%)	10 (12.5%)	57 (71.25%)	3 (3.75%)	3 (3.75%)	6 (7.5%)	80 (100%)	3161.39
Watauga	1 (1.25%)	15 (18.75%)	47 (58.75%)	8 (10.0%)	3 (3.75%)	6 (7.5%)	80 (100%)	3181.14
Wayne	4 (3.2%)	27 (21.6%)	70 (56.0%)	19 (15.2%)	3 (2.4%)	2 (1.6%)	125 (100%)	2562.75
Wilkes	0 (0.0%)	27 (21.6%)	84 (67.2%)	8 (6.4%)	6 (4.8%)	0 (0.0%)	125 (100%)	2141.3
Wilson	7 (5.6%)	26 (20.8%)	77 (61.6%)	8 (6.4%)	5 (4.0%)	2 (1.6%)	125 (100%)	2477.48
Yadkin	2 (1.6%)	16 (12.8%)	92 (73.6%)	9 (7.2%)	4 (3.2%)	2 (1.6%)	125 (100%)	2837.31
Yancey	1 (1.25%)	9 (11.25%)	51 (63.75%)	11 (13.75%)	4 (5.0%)	4 (5.0%)	80 (100%)	3651.11

Table 2: Attributes to be Checked for Attribute Accuracy

Name of Field	Short Description
STREET_NAME	Difference between STREET_NAM and Rte_Nm in the ARCS table
LOC_1_CNTY_CD	Difference between LOC_1_CNTY_CD and the county that ARC is located within based on the COUNTY field based on the Intersect command
ACS_CNTRL_TYP_CD	Indicates some degree of control of through movements to a road. Null indicates that the road does not have any degree of access control. Compared to imagery to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.
TRNLN_LFT_TYP_CD	Represents the type of turning lane. No data indicates that there are no turning lanes present. Compared to imagery to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.
MDN_TYP_CD	Represents the type median present. No data indicates that there is no median present and that the road is not divided. Compared to imagery to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.
MDN_WID	Represents the width of the median. On roads represented as two separate lines (divided), one-half of the median width is stored on each segment. If the road is represented as a single line but has a median (typically because the median length is less than 200 feet), the entire median width is stored on the segment. Negative numbers should be ignored. Median Widths do not contain turn lanes. Compared to imagery to see if field is correct.
NHS_TYP_CD	A network of nationally significant highways approved by Congress in the National Highway System Designation Act of 1995. New routes can also be added to the NHS. No data indicates that the segment is not part of the NHS. All routes on the National Highway System are eligible for federal-aid.Compared to NC_DOT data from FHWA data. Refer to Road_Characteristics_Field_Description for domains.
NBR_LANE_QTY	This represents the through lanes, does not include ancillary lanes used for turning movements and ramps. On divided roads, the value is the number of through lanes in that direction. To estimate for the entire route, double the values on the inventory side. Compared to imagery to see if field is correct.
TRNLN_RGT_TYP_CD	Represents the type of right turning lane. No data indicates that there are no turning lanes present. Compared to imagery to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.

RW_WID	Represents the width of the right of way of the road. Right of Way can vary continuously along the road. The data has been generalized in areas of widely varying Right of Way to represent significant changes. Compared to parcel data to see if field is correct.
SHLDR_WID_LFT_QTY	Represents the total width of the left shoulder. If the Left Shoulder Width is greater than the Left Paved Shoulder Width, then it indicates that a combination shoulder is present, such as bituminous and grass. Values could have 1 decimal point and should be compared to imagery to see if correct.
SHLDR_WID_RGT_QTY	Represents the total width of the right shoulder. If the Right Shoulder Width is greater than the Right Paved Shoulder Width, then it indicates that a combination shoulder is present, such as bituminous and grass. Values could have 1 decimal point and should be compared to imagery to see if correct.
SPD_LMT_TYP_CD	Represents the posted speed limit. If information is not available, an estimate is used. Can be verified through ordnances and PATHWEB for primary highways.
SRFC_TYP_CD	Represents the surface type of the segment. Values are for state-maintained roads, so no data indicates that road may not be state maintained. Compared to imagery, PATHWEB and field surveys to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.
SRFC_WID	Represents the paved surface width, or the road width from ditch to ditch on unpaved roads. The Surface Width does not include the median width. On divided roads, it is the paved width on that side of the median. On paved roads, the Surface Width is edge of pavement to edge of pavement (includes paved shoulders). Compared to imagery to see if field is correct.
TRCK_RTE_TYP_CD	Internal and federally-designated truck routes. No data indicate trucks are allowed on the route without restrictions. Compared to ordnances and shape file to see if correct. Refer to Road_Characteristics_Field_Description for domains.
SW_PVD_LFT_QTY	The paved width of the left shoulder. Are positive numbers with up to 1 decimal point. Compared to imagery to see if field is correct.
STRCTR_CD	Represents if a structure is present. Are sparsely populated. Compared to imagery, PATHWEB and field surveys to see if field is correct. Refer to Road_Characteristics_Field_Description for domains.
SW_PVD_RGT_QTY	The paved width of the right shoulder. Are positive numbers with up to 1 decimal point. Compared to imagery to see if field is correct.
MLTRY_BASE_CD	The military base that the STRAHNET route is located within. Where applicable, but this data item has never been fully populated. Compared to shape file to see if correct. Refer to Road_Characteristics_Field_Description for domains.

SHN_TYP_CD	The military's Strategic Highway Network (a subset of the National Highway System). Where applicable, but this data item has never been fully populated. Compared to shape file to see if correct. Refer to Road_Characteristics_Field_Description for domains.
TRNLN_LFT_WID	The width of the left turning lane. Where applicable, but this data item has never been fully populated. Compared to imagery, PATHWEB and field verification to see if field is correct.
TRNLN_RGT_WID	The width of the right turning lane. Where applicable, but this data item has never been fully populated. Compared to imagery, PATHWEB and field verification to see if field is correct.

Table 3: Attributes Requirements Checked within all Road Characteristics as Part of Attribute Completeness Component

Error #	Field Name	Criteria
1	Beg_Intersect	Must be a populated with a Route Number, Value that begins with 'C', 'DEAD-END', 'X-CROSS' or 'PSEUDO'
2	End_Intersect	Must be a populated with a Route Number, Value that begins with 'C', 'DEAD-END', 'X-CROSS' or 'PSEUDO'
3	Beg_Intersect_Mp	Needs to be populated with a number
4	End_Intersect_Mp	Needs to be populated with a number
5	STREET_NAME	Needs to be populated with a string. Can not be Null or empty
6	Rte_Nm	Needs to be populated with a string. Can not be Null or empty
7	ACS_CNTRL_TYP_CD	Must be 3, 2 or Null
8	TRNLN_LFT_TYP_CD	Must be 0,1,2,3,4,5, or Null
9	MDN_TYP_CD	Must be 1,3,4,5,6,9,10,11 or Null
10	MDN_WID	Needs be populated with a number if MDN_TYP_CD is populated
11	NHS_TYP_CD	Must 0,1,2,3,4,5,6,7,8,9,10,11,12,13 or Null
12	NBR_LANE_QTY	Must be populated with a positive integer. Can not be Null
13	TRNLN_RGT_TYP_CD	Must be 0,1,2,3,4,5 or Null
14	RW_WID	Must be a number. Can not be Null
15	SHLDR_LFT_TYP_CD	Must be 1,2,3,4,5,6. Can not be Null
16	SHLDR_RGT_TYP_CD	Must be 1,2,3,4,5,6. Can not be Null
17	SPD_LMT_TYP_CD	Must be 10,15,20,21,22,25,30,34,35,36,40,45,48,50,51,55,60,65,66,69,70,88. Can not be Null
18	SRFC_TYP_CD	Must be 0,1,2,3,4,5,6,7,8,9,10,11,12,13. Can not be Null
19	SRFC_WID	Must be a positive number. Can not be Null
20	TRCK_RTE_TYP_CD	Must be 2,3,4,5 or Null

21	SW_PVD_LFT_QTY	Must be a positive number or Null
22	STRCTR_CD	Must be 1,2,3 or Null
23	SW_PVD_RGT_QTY	Must be a positive number or Null
24	MLTRY_BASE_CD	Must be 1,2,3,4,5,6,7 or Null
25	SHN_TYP_CD	Must be 1,2 or Null
26	TRNLN_LFT_WID	Must be a positive number or Null
27	TRNLN_RGT_WID	Must be a positive number or Null
28	IMPTYP_CD	Can be NL,NR,NR, RF, RL, RW, RP, RI, MA, MI, CS, BS, RC, AT, SS, RB, IP, UP, 00, 10, 20,30,41,BR, NC, OT, RE, RS, ST or Null
29	IMP_DCMT_TYP_CD	Must be 2,4,5,6,7 or Null
30	IMP_DCMT_ID	Must be a text or Null
31	IMPTYP_DT	Must be a date or Null
32	ADTN_DCMT_TYP_CD	Must be 1,2,3,4 or Null
33	ADTN_DCMT_ID	Must be text or Null
34	ADTN_DT	Must be a date or Null

Table 4: Dependency Requirements Between Different Fields Checked within All Road Characteristics as Part of Attribute Completeness Component

Error #	Error Type
35	MDN_WID must be MDN_TYP co-populated
36	TRNLN_LFT_TYP_CD and TRNLN_LFT_WID must be co-populated
37	TRNLN_RGT_TYP_CD and TRNLN_RGT_WID must be co-populated
38	All IMP_ attributes must be co-populated
39	If ADNT_DT is populated with a legal value (after 12/31/1930), ADTN_DCMT_TYP_CD must be populated
40	ADTN_DCMT_ID and ADTN_DCMT_TYP_CD must be co-populated
41	If ADTN_DT is populated with a legal value (after 12/31/1930), ADTN_DCMT_ID must be populated
42	ADTN_DT must occur after IMPTYP_DT

Table 5: Summary of Horizontal Accuracy by County (in feet)

County	# of Viable Features Checked	Average Horizontal Accuracy	Maximum Horizontal Accuracy	Minimum Horizontal Accuracy	Standard Dev. of Horizontal Accuracy
Alamance	125	4.575674	6.873236	2.586704	2.278801
Alexander	125	2.086904	3.236543	1.13183	1.116243
Alleghany	80	5.16012	8.821947	2.192497	3.518995
Anson	80	1.434094	2.321265	0.686583	0.881965
Ashe	79	6.5507	10.32635	3.135737	3.815629
Avery	80	9.747543	15.60939	5.237375	5.639545
Beaufort	125	8.611139	13.29762	4.503234	4.631186
Bertie	79	7.981504	10.67443	5.315491	2.842595
Bladen	80	7.880269	11.94361	4.422439	3.971831
Brunswick	124	3.720796	5.32358	2.233488	1.629638
Buncombe	198	5.302251	7.876242	3.03314	2.551805
Burke	125	7.635376	15.18756	2.771182	6.917941
Cabarrus	125	0.882381	1.455967	0.354425	0.584216
Caldwell	125	4.993183	7.876506	2.105554	3.095782
Camden	50	6.4667	9.675549	3.760078	3.133055
Carteret	80	5.831258	8.687444	3.466567	2.780159
Caswell	80	7.185082	11.53812	3.109574	4.393305
Catawba	200	1.148711	2.395118	0.257428	1.153987
Chatham	125	1.762807	2.986507	0.754343	1.184983
Cherokee	80	9.46679	12.84978	6.480972	3.298277
Chowan	80	7.296328	9.461163	5.257733	2.208639
Clay	50	18.82125	27.09574	10.33415	8.486332
Cleveland	200	1.131731	1.961221	0.436267	0.812549
Columbus	125	6.176175	9.066839	3.517749	2.90994
Craven	125	16.29051	18.41117	14.31005	2.146751
Cumberland	200	4.964143	7.193488	2.874417	2.269303
Currituck	80	6.372522	9.592303	3.524432	3.157612
Dare	80	5.949038	8.454647	3.756094	2.503961
Davidson	200	0.594621	1.040329	0.22353	0.440911
Davie	125	0.618452	1.019106	0.284954	0.401887

County	# Features	Avg. HA	Max. Acc.	Min. Acc.	STD of Acc.
Duplin	125	5.723279	8.440819	3.206536	2.739801
Durham	125	5.348598	10.36563	1.953291	4.611062
Edgecombe	80	5.361517	7.687056	3.419525	2.285976
Forsyth	200	0.889436	1.541632	0.405467	0.62226
Franklin	125	7.972718	11.37205	4.93251	3.379641
Gaston	200	1.32574	2.498928	0.53565	1.085739
Gates	50	8.014362	11.38695	5.017492	3.363963
Graham	50	5.751653	9.390283	2.723747	3.541725
Granville	125	2.960739	4.732194	1.50239	1.782354
Greene	80	8.278892	15.55458	3.762465	6.614351
Guilford	200	2.527683	4.905994	0.850874	2.201745
Halifax	125	5.924187	8.447715	3.730235	2.439123
Harnett	125	2.259245	3.306019	1.361904	1.024557
Haywood	125	3.970252	6.269165	2.024942	2.231359
Henderson	125	5.641468	8.506676	2.949067	2.94606
Hertford	80	6.520217	9.157847	4.152797	2.623695
Hoke	125	5.96633	7.777576	4.218881	1.853174
Hyde	50	10.1741	14.38706	6.152749	4.215613
Iredell	200	1.375779	2.164925	0.671773	0.780794
Jackson	125	5.655461	9.461264	2.268547	3.790097
Johnston	200	9.323649	12.08316	6.730858	2.820699
Jones	50	2.624797	3.47069	1.792676	0.88045
Lee	125	4.575649	6.845482	2.562252	2.245775
Lenoir	125	7.146384	10.20733	4.240606	3.15647
Lincoln	125	0.92478	1.548427	0.34503	0.650796
Macon	125	10.25005	18.60122	3.608881	7.939877
Madison	80	6.263119	10.09569	2.969445	3.766209
Martin	80	6.704274	8.861238	4.691187	2.219116
McDowell	125	1.780211	2.757278	1.015464	0.936336
Mecklenburg	200	2.384219	3.501488	1.366363	1.134231
Mitchell	80	6.838941	10.32983	3.50481	3.644472
Montgomery	80	6.190773	8.783283	3.773209	2.622428
Moore	125	3.624548	5.896373	1.641185	2.261938
Nash	125	5.963599	8.697408	3.1606	2.931632
New Hanover	125	3.062938	4.551376	1.744833	1.484098

County	# Features	Avg. HA	Max. Acc.	Min. Acc.	STD of Acc.
Northampton	80	6.116438	8.847708	3.654514	2.712072
Onslow	125	5.620067	7.947146	3.373116	2.389015
Orange	125	6.784107	10.02025	3.95838	3.188332
Pamlico	49	2.596698	4.58654	1.058088	1.896576
Pasquotank	79	1.61707	2.366444	0.92512	0.756173
Pender	125	9.81416	13.31729	6.680781	3.457735
Perquimans	80	7.776855	9.739351	5.905481	2.01824
Person	80	2.329803	3.240485	1.41962	0.953948
Pitt	125	7.199033	10.63836	4.077589	3.433833
Polk	80	5.500254	8.844	2.575547	3.27709
Randolph	200	5.692745	8.598025	2.919964	2.893356
Richmond	125	7.585224	9.774725	5.602366	2.19378
Robeson	125	7.136572	10.7564	3.979231	3.570311
Rockingham	125	6.90359	10.1949	4.143026	3.189332
Rowan	200	1.004875	1.611783	0.447341	0.621267
Rutherford	125	5.571318	8.706385	2.826928	3.087882
Sampson	125	11.44374	16.8486	6.771669	5.321635
Scotland	125	0.982493	1.553575	0.483677	0.564139
Stanly	125	5.060093	7.636719	2.701055	2.579671
Stokes	125	0.743669	1.316931	0.275777	0.550738
Surry	125	3.114287	4.997792	1.457683	1.850794
Swain	80	1.173991	2.298579	0.307797	1.06064
Transylvania	80	4.157306	6.519312	1.939963	2.395525
Tyrrell	50	9.215823	12.47811	6.010747	3.419089
Union	200	1.026971	1.62834	0.475578	0.605942
Vance	80	1.792544	2.474216	1.168833	0.691486
Wake	315	2.783344	4.243372	1.443695	1.484958
Warren	80	2.490367	3.572334	1.458668	1.120123
Washington	80	6.539634	9.011218	4.006696	2.643129
Watauga	80	8.816234	11.98564	5.863112	3.214561
Wayne	125	3.194716	4.788875	1.738213	1.613143
Wilkes	125	4.813398	7.709536	2.37849	2.818405
Wilson	125	5.693832	7.606503	3.95086	1.901723
Yadkin	125	9.967822	14.15636	5.974182	4.223526
Yancey	80	6.72495	10.37746	3.484991	3.614184

Table 6: Summary of Attribute Accuracy by County. Accuracy is averaged from the 23 separate attributes checked for accuracy based on the percent of road that was correctly attributed.

COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)	COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)	COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)
Alamance	84.51088977	Forsyth	76.83971104	Onslow	93.94717478
Alexander	88.98092762	Franklin	85.28778445	Orange	82.52475979
Alleghany	91.41879084	Gaston	85.96608882	Pamlico	86.9694659
Anson	80.16269554	Gates	89.9892118	Pasquotank	83.50761748
Ashe	85.92140987	Graham	93.27510798	Pender	95.90718867
Avery	76.70756991	Granville	89.26255312	Perquimans	89.83994901
Beaufort	89.33796926	Greene	75.55602054	Person	89.54913756
Bertie	88.39401004	Guilford	89.18414895	Pitt	85.500874
Bladen	85.42325184	Halifax	85.90605375	Polk	82.70743069
Brunswick	79.79154946	Harnett	78.39894128	Randolph	91.0175861
Buncombe	87.49160749	Haywood	84.25745562	Richmond	93.83198463
Burke	79.12860365	Henderson	89.98492253	Robeson	94.29667683
Cabarrus	77.53000592	Hertford	72.4312672	Rockingham	92.57396392
Caldwell	79.75223221	Hoke	83.18994541	Rowan	92.36723388
Camden	78.801619	Hyde	83.55167018	Rutherford	89.17379652
Carteret	82.84217823	Iredell	88.09049267	Sampson	95.94110486
Caswell	78.38892094	Jackson	94.06792117	Scotland	91.62422188
Catawba	83.22962997	Johnston	90.59821501	Stanly	94.32323225
Chatham	84.74453678	Jones	78.12928842	Stokes	94.66068468
Cherokee	82.41964444	Lee	92.02112485	Surry	87.86498897
Chowan	79.27399229	Lenoir	92.78084594	Swain	94.7885456
Clay	85.80126666	Lincoln	77.07170709	Transylvania	92.28327746
Cleveland	85.04035529	Macon	85.12674412	Tyrrell	95.27677213
Columbus	80.58562462	Madison	86.28357715	Union	94.159066
Craven	78.33328612	Martin	92.0176955	Vance	94.0651231
Cumberland	85.20835434	McDowell	89.15173327	Wake	95.60548742
Currituck	85.11307409	Mecklenburg	92.38123097	Warren	94.21736841
Dare	77.2409003	Mitchell	95.41255805	Washington	88.66494751
Davidson	74.38080625	Montgomery	88.72578698	Watauga	92.92954203

COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)	COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)	COUNTY	Attribute Accuracy (% of Attribute Length Marked as Correct)
Davie	89.95418117	Moore	94.08650786	Wayne	95.18454395
Duplin	83.34609787	Nash	93.51010272	Wilkes	90.56417184
Durham	72.8949883	New Hanover	96.56505834	Wilson	92.48215205
Edgecombe	85.72131766	Northampton	92.70382502	Yadkin	91.41331131
				Yancey	92.0682852

Table 7: Summary of Attribute Accuracy by Attribute. Average is computed based on the average correctness score of all 100 counties.

ATTRIBUTE	AVERAGE	HIGH	LOW
RW_WID	49.23770049	New Hanover County (96.11%)	Caswell County (27.26%)
STRCTR_CD	51.26617557	4 Counties (100%)	Scotland County (1.80%)
TRNLN_RGT_TYP_CD	68.19650727	50 Counties (100%)	6 Counties (0%)
MDN_WID	68.29458407	14 Counties (100%)	Madison County (0%)
MDN_TYP_CD	71.6993525	18 Counties (100%)	Madison County (0%)
TRNLN_LFT_TYP_CD	74.81220855	33 Counties (100%)	Avery, Caswell and Polk Counties (0%)
TRNLN_RGT_WID	76.92659392	39 Counties (100%)	10 Counties (0%)
TRNLN_LFT_WID	79.20348258	37 Counties (100%)	Caswell County (0%)
SRFC_WID	91.04026934	6 Counties (100%)	Durham County (61.17%)
SHLDR_LFT_TYP_CD	94.06483554	23 Counties (100%)	Dare County (69.74%)
SHLDR_RGT_TYP_CD	94.21479728	26 Counties (100%)	Dare County (69.74%)
STREET_NAME	96.70316393	15 Counties (100%)	Martin County (85.07%)
NHS_TYP_CD	97.63297088	42 Counties (100%)	Mecklenburg County (76.80%)
SW_PVD_RGT_QTY	98.08756179	47 Counties (100%)	Chowan County (87.43%)
SRFC_TYP_CD	98.19738182	40 Counties (100%)	Caldwell County (86.81%)
SW_PVD_LFT_QTY	98.27739924	42 Counties (100%)	Chowan County (88.91%)
ACS_CNTRL_TYP_CD	98.90448316	51 Counties (100%)	Durham County (85.46%)
TRCK_RTE_TYP_CD	99.60712625	88 Counties (100%)	Washington County (85.30%)
NBR_LANE_QTY	99.81212264	74 Counties (100%)	Northampton County (96.47%)
LOC_1_CNTY_CD	99.85211551	33 Counties (100%)	Haywood County (97.84%)
SHN_TYP_CD	99.91629562	94 Counties (100%)	Cabarrus County (94.69%)
MLTRY_BASE_CD	100	All Counties (100%)	None
SPD_LMT_TYP_CD	100	All Counties (100%)	None

Table 8: Summary of Attribute and Dependency Errors Contributing to Attribute Completeness Component for all Road Characteristics in County

Error #	Field Name	Criteria	% Rd Chrs Incomplete	High County (% Incomplete)	Low County (% Incomplete)
Attribute Errors					
9	MDN_TYP_CD	Must be 1,3,4,5,6,,9,10,11 or Null	.002%	Lenoir (.22%)	Many counties (0%)
12	NBR_LANE_QTY	Must be populated with a positive integer. Can not be Null	1.7849%	Forsyth (4.21%)	Few counties (0%)
14	RW_WID	Must be a number. Can not be Null	35.93%	Caswell (51.02%)	New Hanover (16.25%)
15	SHLDR_LFT_TYP_CD	Must be 1,2,3,4,5,6. Can not be Null	14.1772%	Mecklenburg (30.40%)	Clay (5.80%)
16	SHLDR_RGT_TYP_CD	Must be 1,2,3,4,5,6. Can not be Null	13.0796%	Mecklenburg (26.77%)	Anson (5.52%)
17	SPD_LMT_TYP_CD	Must be 10, 15, 20, 21, 22, 25, 30, 34, 35, 36, 40, 45, 48, 50, 51, 55, 60, 65, 66, 69, 70, 88. Can not be Null	4.9097%	Chatham (17.85%)	Clay and Hoke (0%)
18	SRFC_TYP_CD	Must be 0,1,2,3,4,5,6,7,8,9,10,11,12,13. Can not be Null	2.9447%	Mecklenburg (9.79%)	Pamlico (.09%)
19	SRFC_WID	Must be a positive number. Can not be Null	3.1633%	Mecklenburg (11.67%)	Pamlico (.09%)
	All Attribute Errors (Map 8)		1.1783%	Mecklenburg (3.87%)	Dare (1.25%)
Dependency Errors					
35	MDN_WID are not MDN_TYP co-populated		.17416%	Pasquotank (.95%)	Many counties (0%)
36	TRNLN_LFT_TYP_CD and TRNLN_LFT_WID are not co-populated		.5789%	Cumberland (2.88%)	Many counties (0%)
37	TRNLN_RGT_TYP_CD and TRNLN_RGT_WID are not co-populated		1.1200%	Tyrrell (5.37%)	Many counties (0%)
38	All IMP_ attributes are not co-populated		86.8488%	Pasquotank (99.89%)	Pamlico (68.26%)
39	ADNT_DT is populated with a legal value (after 12/31/1930), but ADTN_DCMT_TYP_CD is not populated		41.8832%	Dare (62.84%)	Anson (18.40%)
40	ADTN_DCMT_ID and ADTN_DCMT_TYP_CD are not co-populated		.01441%	Craven (.62%)	Many Counties (0%)
41	ADTN_DT is populated with a legal value (after 12/31/1930), but ADTN_DCMT_ID is not populated		41.878%	Dare (62.84%)	Anson (18.40%)
42	ADTN_DT is before IMPTYP_DT		27.3792%	Chowan (42.43%)	Gates (11.19%)
	All Dependency Errors (Map 9)		24.9846%	Dare (33.69%)	Anson (17.11%)
	Attribute Completeness (All Errors as shown in Map 7)		5.7129%	Mecklenburg (8.48%)	Pamlico (4.62%)

Table 9: Summary of Final Score Taken from 3 Metrics Measured. Each Metric (HA, AA, AC) was weighted equally to compute the final score.

County Name	Horizontal Accuracy	Attribute Accuracy	Attribute Completeness	Final Score	County Name	Horizontal Accuracy	Attribute Accuracy	Attribute Completeness	Final Score
Alamance	78.2	50.1	52.2	60.2	Johnston	52.1	75.3	50.1	59.2
Alexander	91.8	68.6	50.5	70.3	Jones	88.9	23.6	73	61.8
Alleghany	75	78.7	59	70.9	Lee	78.2	81.2	44.4	67.9
Anson	95.4	32	98.4	75.3	Lenoir	64.1	84.3	78.5	75.6
Ashe	67.3	55.9	61.6	61.6	Lincoln	98.2	19.2	81	66.1
Avery	49.8	17.7	58	41.8	Macon	47	52.6	48	49.2
Beaufort	56	70.1	60.6	62.2	Madison	68.9	57.4	48.7	58.3
Bertie	59.5	66.1	70.2	65.3	Martin	66.5	81.2	63.2	70.3
Bladen	60	53.8	73.2	62.3	McDowell	93.5	69.3	36.1	66.3
Brunswick	82.8	30.5	32.4	48.6	Mecklenburg	90.2	82.7	0	57.6
Buncombe	74.2	62.4	41.8	59.5	Mitchell	65.7	95.2	64.7	75.2
Burke	61.4	27.8	33.6	40.9	Montgomery	69.3	67.5	56.6	64.5
Cabarrus	98.4	21.1	45.4	55	Moore	83.4	89.7	64	79
Caldwell	75.9	30.3	70.5	58.9	Nash	70.5	87.3	67.6	75.1
Camden	67.8	26.4	61.3	51.8	New Hanover	86.5	100	31	72.5
Carteret	71.3	43.1	44.1	52.8	Northampton	69.7	84	72.8	75.5
Caswell	63.8	24.7	43.5	44	Onslow	72.4	89.2	55	72.2
Catawba	97	44.7	28.4	56.7	Orange	66	41.8	48.2	52
Chatham	93.6	51	23	55.9	Pamlico	89	60.2	100	83.1
Cherokee	51.3	41.4	48.8	47.2	Pasquotank	94.4	45.9	30.3	56.9
Chowan	63.2	28.4	27.8	39.8	Pender	49.4	97.3	55.9	67.5
Clay	0	55.4	79.2	44.9	Perquimans	60.6	72.1	42.1	58.3
Cleveland	97.1	52.2	65.7	71.7	Person	90.5	70.9	60.5	74
Columbus	69.4	33.8	76.1	59.8	Pitt	63.8	54.2	67.5	61.8
Craven	13.9	24.5	40	26.1	Polk	73.1	42.6	51.7	55.8
Cumberland	76	52.9	34.1	54.3	Randolph	72	77	55.1	68
Currituck	68.3	52.5	59.1	60	Richmond	61.6	88.7	48.2	66.2
Dare	70.6	19.9	27.3	39.3	Robeson	64.1	90.6	64.8	73.2
Davidson	100	8.1	40.6	49.6	Rockingham	65.4	83.5	70.8	73.2
Davie	99.9	72.6	62.2	78.2	Rowan	97.7	82.6	72.1	84.1

County Name	Horizontal Accuracy	Attribute Accuracy	Attribute Completeness	Final Score	County Name	Horizontal Accuracy	Attribute Accuracy	Attribute Completeness	Final Score
Duplin	71.9	45.2	73.9	63.7	Rutherford	72.7	69.4	57.6	66.6
Durham	73.9	1.9	42.9	39.6	Sampson	40.5	97.4	86	74.6
Edgecombe	73.8	55.1	72.2	67	Scotland	97.9	79.5	66.9	81.4
Forsyth	98.4	18.3	32.4	49.7	Stanly	75.5	90.7	85.8	84
Franklin	59.5	53.3	80.5	64.4	Stokes	99.2	92.1	73.3	88.2
Gaston	96	56.1	66.6	72.9	Surry	86.2	64	49.7	66.6
Gates	59.3	72.8	82.1	71.4	Swain	96.8	92.6	38.8	76.1
Graham	71.7	86.4	72.1	76.7	Transylvania	80.5	82.3	49.1	70.6
Granville	87	69.7	54.7	70.5	Tyrrell	52.7	94.7	47.1	64.8
Greene	57.8	12.9	93.8	54.8	Union	97.6	90	65.8	84.5
Guilford	89.4	69.4	19.7	59.5	Vance	93.4	89.6	10.3	64.4
Halifax	70.8	55.8	58.4	61.7	Wake	88	96	38.2	74.1
Harnett	90.9	24.7	54.6	56.7	Warren	89.6	90.3	35.3	71.7
Haywood	81.5	49	36.9	55.8	Washington	67.4	67.3	56.5	63.7
Henderson	72.3	72.7	75.2	73.4	Watauga	54.9	84.9	56.4	65.4
Hertford	67.5	0	77.8	48.4	Wayne	85.7	94.3	65.9	82
Hoke	70.5	44.6	61.8	59	Wilkes	76.9	75.1	57.6	69.9
Hyde	47.4	46.1	54.4	49.3	Wilson	72	83.1	59.5	71.5
Iredell	95.7	64.9	50.5	70.4	Yadkin	48.6	78.7	60.9	62.7
Jackson	72.2	89.7	71.1	77.7	Yancey	66.4	81.4	47.9	65.2

Table 10: Breakdown of URBAN vs. RURAL attribute accuracy summarized from county-level measurements.

ATTRIBUTE	*URBAN*	*RURAL*
ACS_CNTRL_TYP_CD	98.50019544	99.28456907
NBR_LANE_QTY	99.82872731	99.85590085
TRNLN_LFT_TYP_CD	82.87595765	74.30945118
TRNLN_LFT_WID	84.68594963	79.94151911
TRNLN_RGT_TYP_CD	84.32103583	70.49262454
TRNLN_RGT_WID	85.81789013	81.37923567
MDN_TYP_CD	82.02494881	69.04220963
MDN_WID	78.20136484	66.72480268

Appendix C: Python Code

Code Sample 1: Python code used to take county name as input parameter and 1) create database with the county name, 2) Clip selected county with Road Characteristics and **LRS_Arcs** 3) Select a random number of **LRS_Arcs** depending upon the number of clipped features in the county 4) Create relationship class between **LRS_Arcs** and Road Characteristics and 5) Export feature classes and relationship class to newly created database.

```
# -*- coding: utf-8 -*-
# ---------------------------------------------------------------------------
# CreateClip.py
# Created on: 2014-03-28 13:13:50.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------------

# Import arcpy module
import arcpy
import arcgisscripting
import os
import sys
import types
import random
import arceditor
from arcpy import env


countyName = arcpy.GetParameterAsText(0).upper()
RP_2014_02_LRS_Arcs = "RP_2014_02_LRS_Arcs_Working"
counties__2_ = "counties"
Counties = "C:\\NC_DOT_Mulrooney\\Data\\Counties"
RP_2014_02_Rd_Char_Mlpst = "RP_2014_02_Rd_Char_Mlpst"

newArcs = countyName + "_LRS_Arcs"
newPost = countyName + "_Rd_Char_Mlpst"
newRandom = countyName + "_RANDOM_LRS_Arcs"

newArcs = "C:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName + "_LRS_Arcs"
newPost = "C:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName + 
"_Rd_Char_Mlpst"
```

```
# Process: Create File GDB
arcpy.CreateFileGDB_management(Counties, countyName, "CURRENT")

# Process: Clip
arcpy.Clip_analysis(RP_2014_02_LRS_Arcs, counties__2_, newArcs, "")

# Process: Clip (2)
arcpy.Clip_analysis(RP_2014_02_Rd_Char_Mlpst, counties__2_, newPost, "")

gp = arcgisscripting.create()
gp.overwriteoutput = 1

#Get the input layer and count the number of features
inputLayer = "C:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName +
"_LRS_Arcs"
outputLayer = "C:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName +
"_RANDOM_LRS_Arcs"

countFeatures = int(arcpy.GetCount_management(inputLayer).getOutput(0))
arcpy.AddMessage("The number of features in " + str(inputLayer) + " is " + str(countFeatures))
arcpy.AddMessage("Output layer is: " + outputLayer)

# determine the number of features that will be checked
if countFeatures >= 10000:
    checkFeatures = 315
elif countFeatures >=3200:
    checkFeatures = 200
elif countFeatures >=1200:
    checkFeatures = 125
elif countFeatures >=500:
    checkFeatures = 80
else:
    checkFeatures = 50

arcpy.AddMessage("The number to be checked in " + str(inputLayer) + " is " + str(checkFeatures))

# Check for the percent of features to be searched
perFeatures = float(100 * float(checkFeatures) / float(countFeatures))
arcpy.AddMessage("The percent features is " + str(perFeatures))

# New Code goes below here
```

```
inputDirName = os.path.dirname(inputLayer)
gp.workspace = inputDirName
desc = gp.describe(inputLayer)
totpnts = gp.getcount(inputLayer)
gp.addmessage("Selecting " + str(checkFeatures) + " random features")

# Generate a list of all features, and select randomly from this
inList = []
randomList = []
fldname = desc.OIDFieldName
rows = gp.SearchCursor(inputLayer)
row = rows.next()
gp.addmessage ("Loading all IDs into a list")

while row:
    id = row.GetValue(fldname)
    inList.append(id)
    row = rows.next()

selpnts = 0
gp.addmessage("Creating the list of randomly selected features")
while len(randomList) < checkFeatures:
    selpnts += 1
    selItem = random.choice(inList)
    randomList.append(selItem)
    inList.remove(selItem)

# Select features whose OID value occurs in the random list, generate
# a *.lyr file from this selection. (Leading and trailing [ and ] marks
# need to be removed from the list object)
theLen = len(str(randomList))
sqlexp = '"' + fldname + '"' + " in " + "(" + str(randomList)[1:theLen - 1] + ")"

# Export the layer to the proper database selected by the user in the parameters
arcpy.Select_analysis(inputLayer, outputLayer, sqlexp)

####  Create Relationshuip

# Local variables:
#ALAMANCE_RANDOM_LRS_Arcs = "ALAMANCE_RANDOM_LRS_Arcs"
#ALAMANCE_Rd_Char_Mlpst = "ALAMANCE_Rd_Char_Mlpst"
```

```
RelateName = "C:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName +
"_Relate"

# Process: Create Relationship Class
arcpy.CreateRelationshipClass_management(newRandom, newPost, RelateName, "SIMPLE", newPost,
newRandom, "BOTH", "ONE_TO_MANY", "NONE", "G1FtSeg", "G1_FtSeg_Id", "", "")
```

Code Sample 2: Python code used to take random **LRS_Arcs** for each county and create 3 equidistant points along arc at the midpoint and then 2 points halfway between the midpoint and the endpoints. NCCU students later digitized points (called centerline points) and compared them to the ‘off’ points.

```
# -*- coding: utf-8 -*-
# ---------------------------------------------------------------------------
# createPoints.py
# Created on: 2014-05-14 09:25:46.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------------

# Set the necessary product code
# import arcinfo


# Import arcpy module
import arcpy

# Get County name from the input parameter
countyName = arcpy.GetParameterAsText(0).upper()

# Load required toolboxes
arcpy.ImportToolbox("C:\Temp\create_points_from_lines\Create Points From
Lines\CreatePointsFromLines.pyt")

# Create Brand New Road Points Layers
## Local variables:
Working_GDB = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb"
centerlinePoints = countyName + "_centerline_Points"

checkDatabase = "T:/NC_DOT_Mulrooney/Data/Counties/" + countyName + ".gdb" + "/" + countyName +
"_centerline_Points"
```

```
if not arcpy.Exists(checkDatabase):
      # Process: Create Feature Class if it does not exist already
      arcpy.CreateFeatureclass_management(Working_GDB, centerlinePoints , "POINT", "", "DISABLED",
"DISABLED",
"PROJCS['NAD_1983_StatePlane_North_Carolina_FIPS_3200_Feet',GEOGCS['GCS_North_American_1983',DATUM['D
_North_American_1983',SPHEROID['GRS_1980',6378137.0,298.257222101]],PRIMEM['Greenwich',0.0],UNIT['Deg
ree',0.0174532925199433]],PROJECTION['Lambert_Conformal_Conic'],PARAMETER['False_Easting',2000000.002
616666],PARAMETER['False_Northing',0.0],PARAMETER['Central_Meridian',-
79.0],PARAMETER['Standard_Parallel_1',34.33333333333334],PARAMETER['Standard_Parallel_2',36.166666666
66666],PARAMETER['Latitude_Of_Origin',33.75],UNIT['Foot_US',0.3048006096012192]];-121841900 -93659000
3048.00609601219;-100000 10000;-100000 10000;3.28083333333333E-03;0.001;0.001;IsHighPrecision", "",
"0", "0", "0")
      arcpy.AddMessage(checkDatabase + " Layer just created")

# Local variables:
RANDOM_LRS_Arcs = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName +
"_RANDOM_LRS_Arcs"

offPoints = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" + countyName +
"_offPoints"
arcpy.AddMessage("Output layer is: " + offPoints)
arcpy.AddMessage("Random Arcs is: " + RANDOM_LRS_Arcs)

# Process: Create Points Along Lines
arcpy.CreatePointsAlongLines_alonglines(RANDOM_LRS_Arcs, offPoints, "0.25", "PERCENTAGE",
"NO_END_POINTS")
```

Code Sample 3: After NCCU students digitized points (called centerline points), a distance was computed between them and the ‘off’ points (points along **LRS_Arcs**). These values were summarized for each **LRS_Arc** and then for each county, and then saved in the working database.

```
# -*- coding: utf-8 -*-
# ---------------------------------------------------------------------------
# createPoints.py
# Created on: 2014-05-14 09:25:46.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------------
```

```
# Set the necessary product code
# import arcinfo


# Import arcpy module
import arcpy
import csv

# Get County name from the input parameter
#countyName = arcpy.GetParameterAsText(0).upper()
countyIndex = 0
countyArray = [None] * 100


with open('C:/Python/countyNames.csv', 'rb') as csvfile:
      dataReader = csv.reader(csvfile, delimiter = ' ', quotechar = '|')

      for row in dataReader:
            countyCand = str(row)
            stringLength = len(str(row)) - 2
            countyArray[countyIndex]=countyCand[2:stringLength].upper()
            arcpy.AddMessage(countyArray[countyIndex])

            countyName = countyArray[countyIndex]
            countyIndex += 1

            countyOffPoints = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" +
countyName + "_offPoints"
            countyCenterPoints = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" +
countyName + "_centerline_Points"

            # Local variables:
            outputTable = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" +
countyName + "_outputTable_NEW"

            # Process: Near
            arcpy.Near_analysis(countyOffPoints, countyCenterPoints, "", "NO_LOCATION", "NO_ANGLE")

            # Process: Summary Statistics
            arcpy.Statistics_analysis(countyOffPoints, outputTable, "NEAR_DIST MEAN;NEAR_DIST
MIN;NEAR_DIST MAX;NEAR_DIST STD", "FID_1")
```

```
            outputStats = "T:\\NC_DOT_Mulrooney\\Data\\Counties\\" + countyName + ".gdb\\" +
countyName + "_statistics_NEW"
            # Summary Statistics
            arcpy.Statistics_analysis(outputTable, outputStats, "MEAN_NEAR_DIST MEAN;MIN_NEAR_DIST
MEAN;MAX_NEAR_DIST MEAN;STD_NEAR_DIST MEAN", "")
```