2018

The South Carolina Safety Belt Study: Large-Scale Location Sampling

Stephanie Jones
University of South Carolina

THE SOUTH CAROLINA SAFETY BELT STUDY: LARGE-SCALE LOCATION SAMPLING

by

Stephanie Jones

Bachelor of Science
University of Denver, 2016

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Statistics

College of Arts and Sciences

University of South Carolina

2018

Accepted by:

John Grego, Director of Thesis

Wilma Sims, Reader

Brian Habing, Reader

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

ABSTRACT

The South Carolina Safety Belt Study is a statewide survey completed yearly to assess the prevalence of safety belt usage on of South Carolina roads through observations from different locations across the state. Every five years the sites for observation are resampled. This thesis breaks down the most recent sampling done for the years of 2018 through 2022. Both the methodology of large scale location sampling and the mathematical idea behind the strategy employed are covered. Further, three different software packages were utilized: R, SAS, and ArcGIS. The steps that were taken and the written function code run for each program is presented. Various roadblocks presented by the data, such as mismatched identification variables, are also described along with their solutions. The final product of the project is a sample of 192 different road reaches in 64 different tracts in 16 different South Carolina counties for counters to go sample from. They are selected through stratified sampling techniques for large scale location data.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Across the country, the past two decades have seen a gradual increase in safety belt usage. The National Highway Traffic Safety Administration (NHTSA) estimates the national seat belt use rate, defined as being restrained by a shoulder-style belt, has jumped from 79.9% in 2000 to 90.1% in 2016 (Li and Pickrell 1). This increase is important, because it has also been accompanied by steady decrease in the percentage of daytime unrestrained passenger vehicle occupant fatalities from 51.6% in 2000 to 40.7% in 2016 (Li and Pickrell 1). This association in conjunction with research that shows a reduction of risk for fatal injury by 45% with the use of a safety belt makes their usage rate a serious focus of state governments to protect their constituents (Li and Pickrell 4).

Accordingly, South Carolina's government has an interest in the state-specific usage rate of safety belts in motorized vehicles by its citizens. The state legislature has made it illegal to not use a proper seat belt by both drivers and passengers when a motor vehicle is in use (South Carolina Seat Belt Law). This makes South Carolina a Primary Law State for safety belt violations, meaning it "gives law enforcement officers the authority to stop a driver if the officer has a clear and unobstructed view of a driver or occupant of a motor vehicle not wearing a safety belt or not secured in a child restraint system" (South Carolina Seat Belt Law). The NHTSA suggests primary law states have consistently higher seat belt usage than secondary law states (Li and Pickrell 1). Despite

the existence of the SC law, many people in South Carolina do not use appropriate safety belts when behind the wheel. A survey conducted by the Department of Statistics in June 2008 found the usage rate for shoulder-style belts to be 79% (Grego 3). Even with the benefit of being a primary law state, this is lower than NHTSA's nation-wide 2008 estimate of 83.1% (Li and Pickrell 1). There is a need for further education and awareness of the benefits of seatbelts in South Carolina until usage rates approach 100%.

Since May of 1991, the office of Highway Safety of the South Carolina Department of Public Safety has created a safety belt awareness drive and commissioned the Statistical Laboratory at the University of South Carolina (Stat Lab) to survey to assess the success of their campaigns (Grego 4). While the commissions were sporadic for a period, they were completed for the years 1991, 1992, 1993, 1994, 1995, 1998, 1999, and every year since there has been a survey before and after the drive to assess the success of it (Grego 4).

To complete this survey individual people who observe motorized vehicles at each location (counters) are sent to specific road locations in the state and record their observations for an hour. For each observation the counter records the type of vehicle, gender of the person, ethnicity of the person, position of the person, and use of seatbelt (Grego 7). Every five years these road locations that are used by the counters are resampled. That requires generating a large-scale location sample that utilizes data from multiple different sources, and it is also the topic of this thesis.

CHAPTER 2

METHODS

The sample of specific road locations must be chosen randomly, but there are ways of randomly sampling which would provide a more desirable random sample than simply selecting 192 road reaches from all of South Carolina. Thus, a couple other factors were considered when designing the sampling in addition to randomness. First, taking the size of the state of South Carolina into consideration, it makes sense that each of the chosen road samples should be relatively close to some group of the other road samples to cut down on travel time. Next, the sample should be a good representation of the different areas of the state as different areas are vastly dissimilar in terms of variables that contribute to safety belt usage. Finally, it must be considered that funding is limited. Therefore, the sample must comprise different representative parts of the state, as sampling the entire state would not feasible within a reasonable budget.

When sampling there are a couple different classic techniques that can be applied while maintaining randomness. A simple random sample without replacement (SRS) is the most basic form of sampling. In SRS every item in the population has an equal probability of being selected (Moore, and Notz 26). A stratified random sample divides the population into groups, called “strata.” From there, a SRS is taken from each of the strata (Moore, and Notz 71). A cluster sample breaks the population down into small groups, called “clusters.” After the clusters are defined, a sample of clusters are then randomly selected and each individual in the cluster is included in the survey (Moore,

and Notz 70). Most complex sampling can be done with these methods or some combination of these methods based on the requirements of the survey. For example, cluster sampling is often used at the last step in location sampling, so the chosen individuals to be surveyed are in close proximity to each other (Moore, and Notz 70). This minimizes the cost of on-location sampling both in time and labor.

Bearing in mind all the previous considerations, a specific sampling scheme was developed by the Stat Lab for this survey. Its refinement in 2003 is still the current method used (Grego 13). Initially, to guarantee the sample is representative, the state is broken down into six different strata. Counties are classified as "upstate," "midlands," or "lowcountry" based on their geographical region (Grego 5). Then each county is classified as either "rural" or "urban," where at least 50% of the population being classified as urban makes the county urban (Grego 5). The combination of these classifications leads to the six strata: upstate rural (UR), upstate urban (UU), midlands rural (MR), midlands urban (MU), lowcountry rural (LR), and lowcountry urban (LU). Then 16 counties are chosen proportionally to their size based on their average daily traffic intensity, where the number of counties per strata sampled is also proportional to their average daily traffic intensity versus the average daily traffic intensity across the state (Grego 5). The variable used to account for these weights was the vehicle miles traveled (VMT). This ensures the sample is representative of traffic across the entire state of South Carolina. Once the counties are selected, four census tracts are randomly sampled from each, again proportionally to their size. This time the proportion is based on their average daily traffic intensity (AADT) (Grego 5). These census tracts are collections of road samples with a geographical proximity to each other, recognizing the

need to cut time and labor cost. From the census tracts, two major road sections and one minor road section were sampled using a SRS. These road sections are where the counters will observe for the survey. The process for the sampling is certainly more complicated than taking a simple random sample, but the stratification captures the considerations of the study while preserving the randomness.

Each road section in South Carolina (that is included in the data for the population for the survey) ends up with a distinct probability of being selected that is the product of the probability of their county being chosen, the conditional probability of their census tract being chosen within each county, and the conditional probability of the specific road section being chosen within each census tract. This "probability of intersection k in tract j in county i in stratum s being included in the multistage sample" can be represented as π_{sijk} (Grego 17). This is just the product of the distinct probability of the county being chosen, the distinct probability of the census tract being chosen, and the distinct probability of the road being chosen. These probabilities are used as inverse weights..

CHAPTER 3

THE DATA

There is not a single data source that appropriately contains the entire population for the sample. Further, the stratification of the population necessitates appropriate data sources to define the strata. Therefore, the data used to complete the sampling came from four original datasets. The original datasets are produced by the specified departments within the state of South Carolina and are publicly available. There was also a dataset created by hand to attach the strata information.

The first dataset is a collection of reaches of the statewide highways in South Carolina as a Shapefile for ArcGIS. Shapefiles include the geographical location of data. These are part of the population of road sections. This dataset is called Statewide Highways. The key variables' first 20 observations are included in Table 3.1. The associated DataBase File (DBF) includes 194,760 rows and 36 variables. Of these 36 variables only certain variables were needed for the project. The variable COUNTY_ID identifies the county in South Carolina with a number between 1 and 46 to correspond with the counties in alphabetical order. The variable ROUTE_TYPE uses character abbreviations to classify the type of route this stretch of the highway is. The variable ROUTE_NUMB is the number associated with the route. The variable ROUTE_DIR represents the direction of the route with the first letter, i.e. "N" for north. The variable BEG_MILEPO is an integer representing the beginning mile post. The variable

END_MILEPO is an integer representing the ending mile post. The variable ROUTE_ID is the name of the highway the reach is a part of, for example "ALLENDALE S-111 N." The variable FEATURE_TY gives more information on which feature of a highway this is, like "I_Ramp." The variable MSLINK is a numeric identification code ranging from 457077 to 865627. The variable STREET_NAM includes the name of the street, i.e. "CROOKED CREEK RD." All of these variables, aside from MSLINK and COUNTY_ID, may or may not be available for each road section. Therefore, there is a lot of missing data.

The second dataset is similar to the statewide highways dataset, but instead includes stretches from non-highway roads. The SC DOT calls these roads "other," and they make up the rest of the population of road sections. This dataset is called Statewide Other. The key variables' first 20 observations are included in Table 3.2. The Shapefile's associated DBF includes 248,547 rows and 36 variables. Most of these variables coincide with the variables in the statewide highway database. The notable exception is that the numeric identification code variable MSLINK is replaced with the variable ID, which ranges from 283167 to 972825. While these variables are slightly different, they both serve the same purpose of identifying the road section. Neither of the key variables in this dataset contain missing values.

The third dataset comes from the Census Bureau, and it contains the census tracts that are used to help rein in the locations of the sample. The key variables' first 20 observations are included in Table 3.3. The Shapefile's associated DBF includes 1103 rows and 11 variables. There are two significant variables in this file. The first is the variable COUNTYFP, which labels the counties in South Carolina alphabetically with

numeric digits 1 through 46. The second is the variable TRACTCE, which attaches the census tract to each observation. There are no missing values in this dataset.

The fourth dataset is a Shapefile of Statewide Traffic Points that has the information for the average daily traffic intensity that is needed for the weights of the samples. The key variables' first 20 observations are included in Table 3.4. The Shapefile's associated DBF includes 11,539 rows and 11 variables. There are four variables that were used. The variable CountyName identifies the county the traffic point is in using the complete name, such as "ABBEVILLE." The variable FactoredAA gives the average daily traffic intensity. The variable Latitude lists the latitude of the traffic point. The variable Longitude lists the longitude of the traffic point. While this file contains detailed information on the traffic intensity, its only location identifying variables are Latitude and Longitude. These variables do not match the specific road reaches from the Statewide Highway and Statewide Other files. Therefore, the combination of datasets are needed. The Statewide Highway and Statewide Other datasets comprise the sampling pool with location information, while the Statewide Traffic Points create the weights. The Census Bureau dataset is needed to match census tracts to the other datasets.

Finally, a small dataset was created by hand to attach strata information to the datasets. It is referred to as ccode29. The key variables' first 20 observations are included in Table 3.5. It has 29 rows and 4 variables. The variable County_ID is a numeric identification of the county. The variable strat is a two-letter abbreviation of the strata this record's county was assigned to, e.g "UU." The variable pop is the population of the county in thousands. The final variable vmt is the total vehicle miles travelled in millions.

There are only 29 counties, as some counties will be removed from the sampling pool, which is explained in the next chapter. This dataset has no missing data and was compiled by information from the South Carolina government.

Each of the Statewide Highway, Statewide Other, and Statewide Traffic Points datasets needed to be joined with the census tracts from the Census Bureau dataset. To do this, they each needed a variable to match on. The Census Bureau had a clear COUNTY variable to match on, so all that was needed was a matching variable in the three datasets. For Statewide Highway and Statewide Other the COUNTY_ID variable was already in a matching format. For Statewide Traffic Points, however, the variable CountyName contained the appropriate information but in a different format. Therefore, some quick SAS code was written to create a variable with this information that was in the same numeric form.

Once the matching variables were determined, a spatial join was preformed to combine each of the three Statewide datasets with their corresponding census tracts. To do so, the software package ArcGIS package was utilized. Each dataset came within a shapefile, which can be opened in a part of the package called ArcMap. Each dataset is mapped onto an interactive map. The underlying observations of a shapefile, which are individually mapped in ArcMap, reside in its attribute table. These shapefiles can be overlaid on each other to see how each data point lines up with another spatially. So, each of the datasets could be visually matched up with a corresponding census tract for the area it is in. Of course manually recording each of these matches would be infeasible, but there is a tool called Spatial Joins built-in to the software that automates this process. ArcMap defines the process as joining “attributes from one feature to another based on

the spatial relationship" (Spatial Join). This completes a one-to-many match by location; within this context, this is one census tract being matched to each of the individual road reaches. The resulting Shapefile from the Spatial Join contains its corresponding attribute table with observations matched. This can be used outside of ArcMap as a DBF. Once exported, the each of the three joins' DBFs which contain the matched information were read into SAS. Ccode29 was also read into SAS to give the VMT for weighting counties, and it was time to begin sampling.

Table 3.1 Statewide Highway Data

COUNTY_ID	ROUTE_TYPE	ROUTE_NUMB	ROUTE_DIR	BEG_MILEPO	END_MILEPO	DATA_SOURC	ROUTE_ID	FEATURE_TY	MSLINK	STREET_NAM
3	S-	275	N	0.05	0.1		ALLENDALE S-275 N	City_Paved_Sec	458601	HILL ST
25	S-	110	E	0.2	0.36	SCC	HAMPTON S-110 E	City_Paved_Sec	459599	MIDDLE ST
25	S-	110	E	0.07	0.14	SCC	HAMPTON S-110 E	City_Paved_Sec	459600	MIDDLE ST
25	S-	110	E	0.36	0.41	SCC	HAMPTON S-110 E	City_Paved_Sec	459601	MIDDLE ST
25	S-	110	E	0.73	0.78	SCC	HAMPTON S-110 E	City_Paved_Sec	459602	WOOD ST
3	S-	112	E	0.24	0.31		ALLENDALE S-112 E	City_Paved_Sec	459603	LEE AVE S
3	S-	112	E	0.18	0.24		ALLENDALE S-112 E	City_Paved_Sec	459604	LEE AVE S
3	S-	112	E	0.12	0.18		ALLENDALE S-112 E	City_Paved_Sec	459605	LEE AVE S
3	S-	112	E	0	0.12		ALLENDALE S-112 E	City_Paved_Sec	459606	LEE AVE N
3	S-	111	N	0	0.07		ALLENDALE S-111 N	City_Paved_Sec	459607	BEAUFORT AVE N
3	S-	111	N	0.07	0.13		ALLENDALE S-111 N	City_Paved_Sec	459608	BEAUFORT AVE N
3	S-	111	N	0.13	0.2		ALLENDALE S-111 N	City_Paved_Sec	459609	BEAUFORT AVE N
3	S-	105	N	0	0.17		ALLENDALE S-105 N	City_Paved_Sec	459610	COTTON ST W
3	S-	105	N	0.17	0.2		ALLENDALE S-105 N	City_Paved_Sec	459611	COTTON ST W
3	S-	105	N	0.2	0.41		ALLENDALE S-105 N	City_Paved_Sec	459612	COTTON ST W
3	S-	106	E	1.21	1.4		ALLENDALE S-106 E	Paved_Sec	459613	BETHEL CHURCH RD
3	S-	106	E	0	1.21		ALLENDALE S-106 E	Paved_Sec	459614	BETHEL CHURCH RD
3	S-	123	N	0	0.06		ALLENDALE S-123 N	City_Paved_Sec	459881	
25	S-	114	E	0.37	0.43	SCC	HAMPTON S-114 E	City_Paved_Sec	459968	THIRD ST
25	S-	114	E	0.43	0.49	SCC	HAMPTON S-114 E	City_Paved_Sec	459969	THIRD ST

Table 3.2 Statewide Other Data

COUNTY_ID	ROUTE_NUMB	STREET_NAM	ROUTE_TYPE	ID	BEG_MILEPO	END_MILEPO	ROUTE_ID	ROUTE_DIR
46	4463	SHADOW LAWN CT	L-	964710	0.178	0.25	L-4463	N
46	4460	SHADE TREE CIR	L-	964708	0	0.244	L-4460	E
26	1030	GRAINLOYD RD	L-	688301	0	0.343	L-1030	E
26	6472	S OAK ST	L-	688302	0	0.071	L-6472	E
26	6472	S OAK ST	L-	688303	0.071	0.138	L-6472	E
26	6472	S OAK ST	L-	688304	0.138	0.203	L-6472	E
26	3374	CHAPIN CIR	L-	688305	0.165	0.376	L-3374	E
1	354	BOWERS RD	L-	284740	0	0.107	L-354	E
26	3374	PINENEEDLE DR	L-	688306	0	0.064	L-3374	E
26	6473	LUMBER ST	L-	688307	0	0.073	L-6473	N
26	6473	LUMBER ST	L-	688308	0.073	0.096	L-6473	N
26	4045	CANAL ST	L-	688309	0	0.025	L-4045	N
26	6474	CLUB CIR	L-	688310	0	0.043	L-6474	E
26	4824	ANTIGUA DR	L-	688311	0.173	0.239	L-4824	E
1	1464	THREE B DR	L-	284747	0	0.241	L-1464	E
46	5409	SHADOW COVE LN	L-	964709	0	0.073	L-5409	N
46	2117	SHADOW LAKES DR	L-	965301	0	0.407	L-2117	N
46	2327	STATEVIEW BLVD	L-	965303	0.055	0.65	L-2327	N
46	1726	TILLMAN ST	L-	965302	0	0.088	L-1726	E
46	654	WATERFORD PARK DR	L-	965304	0.245	0.431	L-654	E

Table 3.3 Census Bureau Data

STATEFP	COUNTYFP	TRACTCE
45	003	020701
45	003	021500
45	007	011001
45	007	012001
45	013	002101
45	013	010900
45	015	020401
45	015	020716
45	019	000200
45	019	000600
45	019	002606
45	019	003600
45	019	004613
45	019	004902
45	019	005100
45	027	960202
45	027	960701
45	031	010100
45	031	011200
45	035	010815

Table 3.4 Statewide Traffic Points Data

CountyName	FactoredAA	ID1	Latitude	Longitude
ABBEVILLE	2300	1	34.193697321731	-82.4014862711541
ABBEVILLE	2500	2	34.196450286258	-82.3973930972611
ABBEVILLE	4800	3	34.4197913120729	-82.3852145519501
ABBEVILLE	5000	4	34.3834395735549	-82.3532441946332
ABBEVILLE	3700	5	34.3709915050999	-82.3376546721275
ABBEVILLE	6600	6	34.1788869315088	-82.3801676355446
ABBEVILLE	2400	7	34.1836865592899	-82.3811426013754
ABBEVILLE	2300	8	34.1858833719841	-82.3809593065369
ABBEVILLE	2100	9	34.1896689411347	-82.3826814283596
ABBEVILLE	2300	10	34.1952658266384	-82.3853382522971
ABBEVILLE	3700	11	34.2209709723427	-82.3895787634777
ABBEVILLE	2200	12	34.2321243575607	-82.3904430145627
ABBEVILLE	1600	13	34.2988494874448	-82.3722809521678
ABBEVILLE	1500	14	34.3262312257558	-82.3782540467296
ABBEVILLE	2000	15	34.3330047879849	-82.3873267589961
ABBEVILLE	1600	16	34.3344286550853	-82.402000901482
ABBEVILLE	750	17	34.3886218878848	-82.4367017598433
ABBEVILLE	1850	18	34.0610125382176	-82.3761504017796
ABBEVILLE	2600	19	34.1425013928865	-82.3879748823615
ABBEVILLE	3900	20	34.1592305626808	-82.4065469519098

Table 3.5 Strata Data

County_ID	Strata	Pop	VMT
2	urban	160.1	4.95
4	urban	187.1	5.69
7	urban	162.2	3.46
8	urban	177.8	4.74
10	urban	350.2	10.3
11	rural	55.3	2.3
12	rural	33.1	1.73
14	rural	35	1.81
15	rural	38.9	2.44
16	rural	68.7	2.1
17	rural	32.1	1.65
18	urban	136.6	3.42
21	urban	136.9	4.58
22	urban	60.2	1.97
23	urban	451.2	11.12
26	urban	269.3	7.45
27	rural	24.8	2.85
28	rural	61.7	2.19
29	urban	76.7	1.88
30	rural	66.5	2.54

CHAPTER 4

SAMPLING

After the Statewide Highway, Statewide Other, and Statewide Traffic Points had the census tracts joined, they were ready to be sampled. First, the Statewide Highway and Statewide Other datasets were joined together in SAS to create one dataset named All that contains the entire population of road reaches to be sampled. In an attempt to create a distinct identification variable in this new dataset, the variable MSLINK from Statewide Highway had 250,000 added to each value. This number was chosen as there were 248,547 observations in the other dataset, so this comes from rounding this number up. This new MSLINK and the Statewide Other variable of ID were combined to make a new identification variable ObjectID. Therefore, the dataset All had the variables COUNTY_ID ROUTE_TYPE ROUTE_NUM ROUTE_DIR BEG_MILEPO END_MILEPO ROUTE_ID FEATURE_TY, STREET_NAM TRACTCE, and the new variable, ObjectID.

Before sampling All had some undesirable observations removed. Due to a request from the National Highway Traffic Safety Association, counties that were deemed to have too few county traffic fatalities were removed using the COUNTY_ID variable. The NHTSA provided a cumulative traffic fatality threshold of 85%, and counties above this were removed. These counties were Abbeville, Allendale, Bamberg, Barnwell, Calhoun, Chesterfield, Edgefield, Fairfield, Greenwood, Hampton, Lee, Marion, Marlboro, McCormick, Newberry, Saluda, and Union. They were removed using

a “not in” statement in SAS. There were also observations that were undesirable to be sampled based on the type of road they came from. Roads that were s-ramps, private, or unpaved were all removed from the sample with “not in” statements in SAS. Additionally, road reaches that did not have a census tract were removed, as they could not be sampled using the current sampling scheme. The last thing done to All was creating a variable roadtype classifying observations with route-type beginning with “L-“ as “Minor,” as these were local roads. All other observations were classified as “Major.”

Before sampling, the dataset from Statewide Highway Traffic Points also needed to be cleaned up. This data was read into a SAS dataset called Tract16. The issues with making Tract16 ready to be sampled were more complex to fix. For each census tract, two major roads and one minor road needed to be available to be sampled. However, many census tracts did not comply with these counts. Specifically, may did not have at least two major roads. To remedy this problem each nonconforming census tract was looked up in ArcGIS and joined to another census tract based on its location. If multiple problem tracts could be combined to meet the requirement, this was given preference. If not, the census tract was matched with the closest census tract that was conforming. Proc Freq and Proc Sql were utilized to identify the problem tracts and exported to Excel where the new assignment of census tracts was recorded. After all revisions were recorded a Data step was used to reassign census tracts within the Tract16 dataset. In one step, census tracts with one major road were identified and merged into a new census tract. In another step, census tracts with zero major roads were identified and merged. This was done in two steps due to confusion as to whether there were any census tracts

with less than one major road from Proc Sql not including observations with values of zero. In the future, this should all be performed in one step.

To sample the counties the ccode29 file had two calculated variables added. The first is sumvmt which collects the sum of the VMT for the strata for each observation. The second is vmtprop which is the proportion of the VMT which is calculated by dividing the individual county's VMT by its strata's sum VMT. This was exported under the name CountyWeights.txt and read into R. In R the counties were selected randomly with respect to their weights as described in Chapter 2. They were then output with their respective probabilities of being selected (computed from their vmtprop). The counties selected were Aiken, Anderson, Charleston, Cherokee, Colleton, Darlington, Dorchester, Greenville, Horry, Lexington, Oconee, Orangeburg, Spartanburg, Sumter, Williamsburg, and York. Then All and Tract16 were narrowed down to only observations in these selected counties.

Similarly, to sample the census tracts, the Tract16 dataset has two calculated variables added. These two variables are meanaadt and wtract, and they are respectively the mean AADT and the proportion of the total AADT. These variables and the counties were exported under the name TractWeights.txt and again read into R. Here 64 census tracts are sampled with four coming from each of the counties in the sample frame. The sampling is done the same way as the counties, except there are two added steps before exporting. First the county probabilities are added to each of the census tracts. Second the probability of the census tract being selected and the probability of the county being selected are multiplied to get each tract's total probability of being chosen. The sampling

and calculations steps are completed three times, so there are backup samples in case the counters are unable to survey in a location.

To finish up the road sections were sampled. To do so R is again utilized. First some cleaning must be done in R to run the functions: the census tract codes were padded with zeros, as some county codes were shortened by a single digit, and the county and census tract probabilities needed to be rounded. Sampling pools of Major roads and Minor roads also had to be made for each of the sampled census tracts. After this a simple random sample of one Minor road and two Major roads from each census tract was taken, giving 192 total distinct road reaches. The probability of each road sample within its census tract was calculated as one divided by the total number of either Major or Minor roads in the respective census tract. This probability was also multiplied by the county and census tract probabilities to get the total probability of the sampled road reaches. This is exported as RoadSamp.txt and read into SAS. There are also two other copies of this sampling done in case the counters encounter bad locations in the sample.

These sampled road reaches are read back into SAS where they are matched with the All dataset to get the location information needed to give to the counters. This match is done with the variables RoadSamp and ObjectID. After this match is done, however, an issue became obvious: some roads were being matched to the wrong road from the All dataset. The reason for this turns out to be overlapping ID variable values in the Statewide Highway's MSLINK and the Statewide Other's ID variable, from ID not starting at one as had been assumed. This is fixed with manual individual lookups, as it is clear which observation is the correct match based on county and tract. Nevertheless, this is something that should be fixed moving forward. In the future, spatial joins create their

own “ObjectID” variable for each record. This would be distinct and sure to start from one. This should have been utilized for identifying Statewide Other’s records, and this same ObjectID variable, plus 250,000, should have been utilized for identifying Statewide Highway’s records. Once this was corrected, however, the rest of matching went smoothly, and location data was added to the sample.

Nevertheless, the sample was not complete. Looking at the final sample, there were three records that required resampling. All three were labeled as “Major roads,” but were not something that would reasonably be referred to as such. Two of these records came from the same census tract and were tiny connector roads near a retail complex. Originally these two roads were just going to be resampled from the census tract, but the census tract only had two major roads. Instead a whole new census tract was sampled and the minor road from the old census tract was removed to be resampled from the new census tract as well. Similarly, the third road sample came from a census tract that only had two major roads. This time it was one that had been built from tracts with not enough major roads. A whole new census tract and three road reaches were sampled. To avoid this problem in the future the minimum number of major roads may be raised. However, doing so would increase the number of individual census tract lookups in ArcGIS which is a lengthy process. Nevertheless, the sample was finally successful and ready to be sent to the state for approval.

CHAPTER 5

RESULTS

The final sample is featured in Tables 5.1 to 5.16 separated by county. It has a total of 192 road reaches, 128 major roads, 64 minor roads, 64 distinct census tracts, and 16 distinct counties.

The counters need to be able to find the exact location they should be surveying through specific directions. This presented a problem. The biggest obstacle was that the latitude and longitude from the Traffic Points dataset were too general. After matching each road sample with the latitude and longitude for the final sample, it becomes apparent many have the same values listed within tracts. The data for the latitude and longitude was likely pulled from a general database, and it would not be enough information to identify where the actual road sample is. Also, some roads had missing data for street name and route type. There was not appropriate information in any of the datasets to identify where all the sections of the road were. Therefore, a feature of ArcGIS had to be utilized to locate each of the road samples. By highlighting and identifying a record in a Shapefile, ArcMap can produce the Degrees Minutes Seconds (DMS) coordinates. After going through the 192 road samples and recording the DMS coordinates, they can be entered into google maps. Google maps produces a map of the area, and clicking on the interactive map gives the city and zipcode of an address that is negligibly close. Maps are created with this information and the DMS coordinates, so that counters can enter either

in their personal GPS and feel confident that they are arriving at the correct location. An example of these maps is included in Appendix E.

In April 2018 a smaller scale pre-survey will be completed with the new locations. Only six counties will be surveyed to keep costs down. To ensure this sample is still representative, each strata was sampled, one county was selected from each of the strata: UU, UR, MU, MR, LU, LR. To do this sampling SRS was used. This is appropriate as the 16 counties that make up the populations of each strata have already been chosen based on their weight. This sample is featured in Table 5.17.

Table 5.1 Aiken Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
MU	2	Aiken	2020400	921916	Minor	0.148916968	0.07996002	0.00662	0.000079	81°51'8.046"W 33°31'49.528"N
MU	2	Aiken	2020601	921992	Minor	0.148916968	0.06405694	0.00855	0.000082	81°53'59.727"W 33°34'30.601"N
MU	2	Aiken	2021203	923049	Major	0.148916968	0.151612903	0.5	0.011289	80°35'10.035"W 33°13'27.581"N
MU	2	Aiken	2021203	923070	Major	0.148916968	0.151612903	0.5	0.011289	80°34'32.198"W 33°12'19.085"N
MU	2	Aiken	2021201	923077	Major	0.148916968	0.148473636	0.25	0.005528	80°33'25.355"W 33°11'53.57"N
MU	2	Aiken	2021201	923110	Minor	0.148916968	0.148473636	0.00319	0.000071	81°44'43.071"W 33°30'7.423"N
MU	2	Aiken	2021203	924257	Minor	0.148916968	0.151612903	0.00308	0.000069	81°43'34.945"W 33°31'3.974"N
MU	2	Aiken	2020601	1067247	Major	0.148916968	0.06405694	0.22222	0.00212	81°55'24.298"W 33°31'49.335"N
MU	2	Aiken	2020400	1070525	Major	0.148916968	0.07996002	0.22222	0.002646	81°50'34.051"W 33°36'17.238"N
MU	2	Aiken	2020400	1070542	Major	0.148916968	0.07996002	0.22222	0.002646	81°50'30.417"W 33°36'24.257"N
MU	2	Aiken	2021201	1070799	Major	0.148916968	0.148473636	0.25	0.005528	81°46'44.359"W 33°33'41.224"N
MU	2	Aiken	2020601	1074506	Major	0.148916968	0.06405694	0.22222	0.00212	81°55'36.016"W 33°33'5.415"N

Table 5.2 Anderson Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UU	4	Anderson	4000200	578798	Minor	0.166179907	0.009487666	0.01667	0.000026	82°38'8.118"W 34°31'2.116"N
UU	4	Anderson	4011401	589350	Minor	0.166179907	0.029739777	0.00485	0.000024	82°29'54.881"W 34°31'41.467"N
UU	4	Anderson	4010102	589654	Minor	0.166179907	0.052499302	0.00285	0.000025	82°27'47.68"W 34°47'33.794"N
UU	4	Anderson	4000200	819894	Major	0.166179907	0.009487666	0.4	0.000631	82°38'54.746"W 34°31'48.163"N
UU	4	Anderson	4000200	819995	Major	0.166179907	0.009487666	0.4	0.000631	82°38'47.583"W 34°31'48.083"N
UU	4	Anderson	4011401	820073	Major	0.166179907	0.029739777	1	0.004942	82°30'20.732"W 34°31'5.839"N
UU	4	Anderson	4010102	821852	Major	0.166179907	0.052499302	0.08	0.000698	82°28'24.755"W 34°45'43.047"N
UU	4	Anderson	4010102	824882	Major	0.166179907	0.052499302	0.08	0.000698	82°29'19.139"W 34°43'56.543"N
UU	4	Anderson	4011401	833652	Major	0.166179907	0.029739777	1	0.004942	82°29'9.169"W 34°30'57.475"N
UU	4	Anderson	4010300	581967	Minor	0.166179907	0.038432891	0.00397	0.00003	82°28'54.654"W 34°39'48.085"N
UU	4	Anderson	4010300	822505	Major	0.166179907	0.038432891	0.33333	0.00213	82°28'45.438"W 34°38'3.663"N
UU	4	Anderson	4010300	820178	Major	0.166179907	0.038432891	0.33333	0.00213	82°30'46.493"W 34°39'52.009"N

Table 5.3 Charleston Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
LU	10	Charleston	10004606	786464	Minor	0.328653478	0.083946188	0.00226	0.000062	79°52'50.818"W 32°49'3.447"N
LU	10	Charleston	10003104	787705	Minor	0.328653478	0.010271318	0.02439	0.000082	80°2'18.017"W 32°55'25.439"N
LU	10	Charleston	10004613	789767	Minor	0.328653478	0.013902298	0.01515	0.000069	79°51'55.872"W 32°50'5.07"N
LU	10	Charleston	10004608	791442	Minor	0.328653478	0.138785835	0.00124	0.000057	79°49'46.688"W 32°53'43.269"N
LU	10	Charleston	10003104	945235	Major	0.328653478	0.010271318	0.16667	0.000563	80°1'53.12"W 32°55'23.003"N
LU	10	Charleston	10003104	945339	Major	0.328653478	0.010271318	0.16667	0.000563	80°2'7.744"W 32°55'47.981"N
LU	10	Charleston	10004613	946830	Major	0.328653478	0.013902298	0.33333	0.001523	79°51'55.725"W 32°50'26.201"N
LU	10	Charleston	10004606	947161	Major	0.328653478	0.083946188	0.08	0.002207	79°50'58.706"W 32°48'51.331"N
LU	10	Charleston	10004613	947249	Major	0.328653478	0.013902298	0.33333	0.001523	79°51'37.175"W 32°50'17.993"N
LU	10	Charleston	10004608	947808	Major	0.328653478	0.138785835	0.10526	0.004801	79°46'35.208"W 32°52'18.712"N
LU	10	Charleston	10004608	947956	Major	0.328653478	0.138785835	0.10526	0.004801	79°42'46.553"W 32°55'58.299"N
LU	10	Charleston	10004606	1074826	Major	0.328653478	0.083946188	0.08	0.002207	79°50'36.896"W 32°49'9.116"N

Table 5.4 Cherokee Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UR	11	Cherokee	11970301	842683	Minor	0.273809524	0.203539823	0.00758	0.000422	81°41'16.074"W 35°4'59.637"N
UR	11	Cherokee	11970202	844162	Minor	0.273809524	0.227467811	0.00694	0.000433	81°35'32.161"W 35°7'19.447"N
UR	11	Cherokee	11970401	844366	Minor	0.273809524	0.179331307	0.0122	0.000599	81°30'46.176"W 35°7'44.997"N
UR	11	Cherokee	11970402	844517	Minor	0.273809524	0.191616767	0.00877	0.00046	81°29'57.661"W 35°7'21.026"N
UR	11	Cherokee	11970301	1015413	Major	0.273809524	0.203539823	0.33333	0.018577	81°41'22.507"W 35°5'29.703"N
UR	11	Cherokee	11970202	1015451	Major	0.273809524	0.227467811	0.13333	0.008304	81°40'2.892"W 35°5'49.587"N
UR	11	Cherokee	11970301	1015452	Major	0.273809524	0.203539823	0.33333	0.018577	81°40'0.175"W 35°5'47.992"N
UR	11	Cherokee	11970202	1016824	Major	0.273809524	0.227467811	0.13333	0.008304	81°36'4.022"W 35°6'37.632"N
UR	11	Cherokee	11970401	1017369	Major	0.273809524	0.179331307	0.05556	0.002728	81°28'49.377"W 35°9'24.982"N
UR	11	Cherokee	11970401	1017373	Major	0.273809524	0.179331307	0.05556	0.002728	81°28'43.621"W 35°9'28.183"N
UR	11	Cherokee	11970402	1017445	Major	0.273809524	0.191616767	0.14286	0.007495	81°26'59.498"W 35°9'54.264"N
UR	11	Cherokee	11970402	1017451	Major	0.273809524	0.191616767	0.14286	0.007495	81°26'45.577"W 35°10'2.126"N

Table 5.5 Colleton Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
LR	15	Colleton	15970300	775099	Minor	0.300492611	0.357541899	0.04167	0.004477	80°43'26.275"W 32°53'22.304"N
LR	15	Colleton	15970800	775755	Minor	0.300492611	0.449760766	0.01136	0.001536	80°19'54.54"W 32°29'3.832"N
LR	15	Colleton	15970600	776269	Minor	0.300492611	0.467592593	0.01099	0.001544	80°37'53.609"W 32°56'6.507"N
LR	15	Colleton	15970402	776336	Minor	0.300492611	0.33908046	0.02174	0.002215	80°38'28.429"W 32°58'15.571"N
LR	15	Colleton	15970300	935990	Major	0.300492611	0.357541899	0.05	0.005372	80°42'53.277"W 32°52'46.8"N
LR	15	Colleton	15970300	936017	Major	0.300492611	0.357541899	0.05	0.005372	80°42'7.048"W 32°54'22.89"N
LR	15	Colleton	15970402	936256	Major	0.300492611	0.33908046	0.15385	0.015676	80°41'18.03"W 32°55'46.803"N
LR	15	Colleton	15970402	936257	Major	0.300492611	0.33908046	0.15385	0.015676	80°41'20.297"W 32°55'47.305"N
LR	15	Colleton	15970800	936668	Major	0.300492611	0.449760766	0.33333	0.04505	80°37'9.24"W 32°43'29.855"N
LR	15	Colleton	15970800	936818	Major	0.300492611	0.449760766	0.33333	0.04505	80°32'52.027"W 32°44'52.526"N
LR	15	Colleton	15970600	936829	Major	0.300492611	0.467592593	0.2	0.028102	80°33'31.899"W 32°49'28.679"N
LR	15	Colleton	15970600	937712	Major	0.300492611	0.467592593	0.2	0.028102	80°38'21.552"W 32°54'35.718"N

Table 5.6 Darlington Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
MR	16	Darlington	16010400	804221	Minor	0.202702703	0.152212389	0.0122	0.000376	80°7'8.767"W 34°21'39.342"N
MR	16	Darlington	16010600	804537	Minor	0.202702703	0.166956522	0.01075	0.000364	80°2'46.578"W 34°21'46.384"N
MR	16	Darlington	16010700	804667	Minor	0.202702703	0.069902913	0.02941	0.000417	80°4'24.077"W 34°22'0.331"N
MR	16	Darlington	16011300	805858	Minor	0.202702703	0.130671506	0.01449	0.000384	79°46'17.74"W 34°15'18.098"N
MR	16	Darlington	16010400	966803	Major	0.202702703	0.152212389	0.5	0.015427	80°6'37.843"W 34°21'31.751"N
MR	16	Darlington	16010400	967156	Major	0.202702703	0.152212389	0.5	0.015427	80°7'47.737"W 34°22'0.857"N
MR	16	Darlington	16010600	967534	Major	0.202702703	0.166956522	0.66667	0.022562	80°4'4.204"W 34°22'35.081"N
MR	16	Darlington	16010700	967809	Major	0.202702703	0.069902913	1	0.01417	80°3'36.287"W 34°21'20.291"N
MR	16	Darlington	16010700	967810	Major	0.202702703	0.069902913	1	0.01417	80°3'34.386"W 34°21'17.575"N
MR	16	Darlington	16010600	967817	Major	0.202702703	0.166956522	0.66667	0.022562	80°3'26.157"W 34°21'26.306"N
MR	16	Darlington	16011300	969596	Major	0.202702703	0.130671506	0.66667	0.017658	79°49'37.606"W 34°15'7.488"N
MR	16	Darlington	16011300	969797	Major	0.202702703	0.130671506	0.66667	0.017658	79°45'23.217"W 34°15'16.352"N

Table 5.7 Dorchester Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
LU	18	Dorchester	18010505	741529	Minor	0.109125718	0.275139665	0.00513	0.000154	80°12'36.97"W 32°58'21.498"N
LU	18	Dorchester	18010504	741553	Minor	0.109125718	0.21719457	0.00714	0.000169	80°13'46.154"W 32°59'24.359"N
LU	18	Dorchester	18010300	923284	Major	0.109125718	0.143373494	0.04167	0.00065	80°30'56.408"W 33°17'16.281"N
LU	18	Dorchester	18010300	923229	Major	0.109125718	0.143373494	0.04167	0.00065	80°33'33.502"W 33°17'31.351"N
LU	18	Dorchester	18010300	744830	Minor	0.109125718	0.143373494	0.01408	0.00022	80°23'23.664"W 33°8'23.334"N
LU	18	Dorchester	18010604	743362	Minor	0.109125718	0.199074074	0.00787	0.000171	80°12'25.721"W 33°1'15.968"N
LU	18	Dorchester	18010504	923956	Major	0.109125718	0.21719457	0.5	0.011851	80°14'59.56"W 32°58'21.777"N
LU	18	Dorchester	18010504	923959	Major	0.109125718	0.21719457	0.5	0.011851	80°14'58.85"W 32°58'22.188"N
LU	18	Dorchester	18010505	924142	Major	0.109125718	0.275139665	1	0.030025	80°13'0.948"W 32°59'50.642"N
LU	18	Dorchester	18010505	924144	Major	0.109125718	0.275139665	1	0.030025	80°12'59.574"W 32°59'51.419"N
LU	18	Dorchester	18010604	924897	Major	0.109125718	0.199074074	1	0.021724	80°11'54.276"W 33°1'59.942"N
LU	18	Dorchester	18010604	924898	Major	0.109125718	0.199074074	1	0.021724	80°11'55.528"W 33°2'1.057"N

Table 5.8 Greenville Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UU	23	Greenville	23003601	890823	Minor	0.324766355	0.026830492	0.00495	0.000043	82°28'39.174"W 34°49'47.832"N
UU	23	Greenville	23003701	893325	Minor	0.324766355	0.029021766	0.00437	0.000041	82°28'2.296"W 34°54'32.44"N
UU	23	Greenville	23003101	903647	Minor	0.324766355	0.027805611	0.00513	0.000046	82°11'23.018"W 34°42'11.245"N
UU	23	Greenville	23000100	904888	Minor	0.324766355	0.016846105	0.00806	0.000044	82°23'9.441"W 34°51'43.24"N
UU	23	Greenville	23003601	1047609	Major	0.324766355	0.026830492	0.16667	0.001452	82°27'46.768"W 34°50'1.118"N
UU	23	Greenville	23003601	1047685	Major	0.324766355	0.026830492	0.16667	0.001452	82°26'55.063"W 34°49'59.591"N
UU	23	Greenville	23003701	1048912	Major	0.324766355	0.029021766	0.66667	0.006284	82°27'44.335"W 34°54'54.28"N
UU	23	Greenville	23003701	1048916	Major	0.324766355	0.029021766	0.66667	0.006284	82°27'44.502"W 34°54'55.233"N
UU	23	Greenville	23003101	1052532	Major	0.324766355	0.027805611	0.07407	0.000669	82°13'10.587"W 34°41'21.954"N
UU	23	Greenville	23003101	1052536	Major	0.324766355	0.027805611	0.07407	0.000669	82°13'8.621"W 34°41'22.248"N
UU	23	Greenville	23000100	1053667	Major	0.324766355	0.016846105	0.22222	0.001216	82°23'7.857"W 34°51'11.917"N
UU	23	Greenville	23000100	1053675	Major	0.324766355	0.016846105	0.22222	0.001216	82°23'6.888"W 34°51'11.243"N

Table 5.9 Horry Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
LU	26	Horry	26060204	685134	Minor	0.23771538	0.046062053	0.00546	0.00006	78°57'17.551"W 33°43'6.452"N
LU	26	Horry	26060203	691443	Minor	0.23771538	0.050593824	0.00505	0.000061	78°57'41.533"W 33°43'58.932"N
LU	26	Horry	26051700	693637	Minor	0.23771538	0.042634731	0.00588	0.00006	78°55'21.129"W 33°39'26.127"N
LU	26	Horry	26020100	698251	Minor	0.23771538	0.007203179	0.03846	0.000066	78°57'35.916"W 34°12'7.173"N
LU	26	Horry	26060203	903398	Major	0.23771538	0.050593824	0.13333	0.001604	78°56'43.124"W 33°43'58.371"N
LU	26	Horry	26020100	903425	Major	0.23771538	0.007203179	0.66667	0.001142	78°58'33.594"W 34°7'35.624"N
LU	26	Horry	26051700	903443	Major	0.23771538	0.042634731	0.25	0.002534	78°57'35.135"W 33°37'53.394"N
LU	26	Horry	26060203	903445	Major	0.23771538	0.050593824	0.13333	0.001604	78°57'0.974"W 33°44'18.738"N
LU	26	Horry	26020100	903457	Major	0.23771538	0.007203179	0.66667	0.001142	78°56'28.469"W 34°6'31.785"N
LU	26	Horry	26060204	903533	Major	0.23771538	0.046062053	0.2	0.00219	78°56'44.428"W 33°43'56.618"N
LU	26	Horry	26051700	903537	Major	0.23771538	0.042634731	0.25	0.002534	78°57'14.307"W 33°38'11.611"N
LU	26	Horry	26060204	903583	Major	0.23771538	0.046062053	0.2	0.00219	78°55'50.132"W 33°43'12.442"N

Table 5.10 Lexington Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
MU	32	Lexington	32020703	452036	Minor	0.251504212	0.02945818	0.00917	0.000068	81°9'3.992"W 33°56'5.024"N
MU	32	Lexington	32021030	609223	Minor	0.251504212	0.035546262	0.008	0.000072	81°12'36.608"W 33°58'56.039"N
MU	32	Lexington	32020507	613197	Minor	0.251504212	0.052631579	0.00503	0.000067	81°8'20.026"W 33°59'39.18"N
MU	32	Lexington	32021028	613324	Minor	0.251504212	0.035798275	0.00769	0.000069	81°16'35.268"W 33°58'50.29"N
MU	32	Lexington	32020507	780223	Major	0.251504212	0.052631579	0.33333	0.004412	81°6'29.653"W 33°58'53.21"N
MU	32	Lexington	32021028	780323	Major	0.251504212	0.035798275	0.28571	0.002572	81°15'23.829"W 33°59'1.972"N
MU	32	Lexington	32020703	781152	Major	0.251504212	0.02945818	0.66667	0.004939	81°8'6.411"W 33°54'40.056"N
MU	32	Lexington	32020703	781153	Major	0.251504212	0.02945818	0.66667	0.004939	81°8'8.866"W 33°54'37.865"N
MU	32	Lexington	32021028	793804	Major	0.251504212	0.035798275	0.28571	0.002572	81°14'0.976"W 33°59'21.496"N
MU	32	Lexington	32021030	794882	Major	0.251504212	0.035546262	0.18182	0.001625	81°13'52.533"W 33°57'33.284"N
MU	32	Lexington	32021030	794885	Major	0.251504212	0.035546262	0.18182	0.001625	81°11'40.874"W 33°58'49.687"N
MU	32	Lexington	32020507	794981	Major	0.251504212	0.052631579	0.33333	0.004412	81°6'26.57"W 33°59'21.638"N

Table 5.11 Oconee Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UR	37	Oconee	37030402	877851	Minor	0.217857143	0.06610338	0.00826	0.000119	83°2'21.272"W 34°44'48.191"N
UR	37	Oconee	37030800	878152	Minor	0.217857143	0.192175408	0.00226	0.000095	82°59'34.542"W 34°41'13.258"N
UR	37	Oconee	37030500	878542	Minor	0.217857143	0.150158299	0.00309	0.000101	82°59'52.922"W 34°42'28.107"N
UR	37	Oconee	37030702	880721	Minor	0.217857143	0.138468592	0.00338	0.000102	82°54'33.949"W 34°41'1.92"N
UR	37	Oconee	37030402	1036966	Major	0.217857143	0.06610338	0.16667	0.0024	83°2'50.636"W 34°45'10.614"N
UR	37	Oconee	37030402	1036970	Major	0.217857143	0.06610338	0.16667	0.0024	83°2'43.464"W 34°45'5.358"N
UR	37	Oconee	37030500	1037029	Major	0.217857143	0.150158299	0.25	0.008178	83°1'46.726"W 34°44'6.555"N
UR	37	Oconee	37030500	1037034	Major	0.217857143	0.150158299	0.25	0.008178	83°1'48.041"W 34°44'9.208"N
UR	37	Oconee	37030800	1037152	Major	0.217857143	0.192175408	0.5	0.020933	83°0'23.96"W 34°41'0.657"N
UR	37	Oconee	37030800	1037513	Major	0.217857143	0.192175408	0.5	0.020933	82°58'50.584"W 34°41'35.565"N
UR	37	Oconee	37030702	1038669	Major	0.217857143	0.138468592	0.33333	0.010055	82°51'51.623"W 34°41'24.84"N
UR	37	Oconee	37030702	1038674	Major	0.217857143	0.138468592	0.33333	0.010055	82°51'50.246"W 34°41'23.335"N

Table 5.12 Orangeburg Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
MR	38	Orangeburg	38011200	817555	Minor	0.426640927	0.113323124	0.01408	0.000681	80°51'7.354"W 33°29'45.794"N
MR	38	Orangeburg	38011000	817893	Minor	0.426640927	0.11196319	0.01429	0.000682	80°52'32.437"W 33°31'58.18"N
MR	38	Orangeburg	38010800	818138	Minor	0.426640927	0.064620355	0.06667	0.001838	80°48'34.914"W 33°32'43.456"N
MR	38	Orangeburg	38010300	819808	Minor	0.426640927	0.203576341	0.01087	0.000944	80°23'53.609"W 33°26'36.937"N
MR	38	Orangeburg	38011200	991206	Major	0.426640927	0.113323124	0.66667	0.032232	80°50'28.236"W 33°29'52.728"N
MR	38	Orangeburg	38011200	991427	Major	0.426640927	0.113323124	0.66667	0.032232	80°49'5.949"W 33°30'21.528"N
MR	38	Orangeburg	38011000	991706	Major	0.426640927	0.11196319	0.66667	0.031845	80°52'26.254"W 33°31'14.075"N
MR	38	Orangeburg	38011000	992280	Major	0.426640927	0.11196319	0.66667	0.031845	80°51'0.858"W 33°32'4.474"N
MR	38	Orangeburg	38010800	992540	Major	0.426640927	0.064620355	0.08	0.002206	80°49'33.346"W 33°33'38.637"N
MR	38	Orangeburg	38010800	992541	Major	0.426640927	0.064620355	0.08	0.002206	80°49'30.753"W 33°33'36.705"N
MR	38	Orangeburg	38010300	993376	Major	0.426640927	0.203576341	0.03571	0.003102	80°31'11.411"W 33°22'56.499"N
MR	38	Orangeburg	38010300	993859	Major	0.426640927	0.203576341	0.03571	0.003102	80°29'54.513"W 33°24'46.981"N

Table 5.13 Spartanburg Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UU	42	Spartanburg	42022500	859099	Minor	0.263434579	0.035494558	0.00355	0.000033	82°7'35.742"W 35°7'23.775"N
UU	42	Spartanburg	42023002	859601	Minor	0.263434579	0.019838884	0.00671	0.000035	82°3'53.181"W 34°56'32.997"N
UU	42	Spartanburg	42021200	865782	Major	0.263434579	0.030562493	1	0.008051	81°20'48.938"W 34°21'12.642"N
UU	42	Spartanburg	42021200	865913	Major	0.263434579	0.030562493	1	0.008051	80°53'50.318"W 34°27'34.562"N
UU	42	Spartanburg	42021200	867138	Minor	0.263434579	0.030562493	0.00392	0.000032	81°54'25.025"W 34°56'57.352"N
UU	42	Spartanburg	42023900	870141	Minor	0.263434579	0.035037879	0.00345	0.000032	81°46'47.032"W 34°54'47.033"N
UU	42	Spartanburg	42022500	1026494	Major	0.263434579	0.035494558	0.11111	0.001039	82°9'31.904"W 35°11'4.37"N
UU	42	Spartanburg	42022500	1026519	Major	0.263434579	0.035494558	0.11111	0.001039	82°8'1.384"W 35°10'2.568"N
UU	42	Spartanburg	42023002	1026747	Major	0.263434579	0.019838884	0.125	0.000653	82°3'15.563"W 34°57'41.177"N
UU	42	Spartanburg	42023002	1026756	Major	0.263434579	0.019838884	0.125	0.000653	82°3'8.442"W 34°57'36.176"N
UU	42	Spartanburg	42023900	1031265	Major	0.263434579	0.035037879	0.33333	0.003077	81°47'44.423"W 34°54'18.611"N
UU	42	Spartanburg	42023900	1031267	Major	0.263434579	0.035037879	0.33333	0.003077	81°47'45.275"W 34°54'19.34"N

Table 5.14 Sumter Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
MU	43	Sumter	43000800	969727	Minor	0.082129964	0.037100949	0.03571	0.000109	80°21'11.409"W 33°56'51.918"N
MU	43	Sumter	43001802	970012	Minor	0.082129964	0.17820324	0.00463	0.000068	80°25'1.686"W 33°57'14.271"N
MU	43	Sumter	43000202	970866	Minor	0.082129964	0.134883721	0.00602	0.000067	80°26'27.559"W 33°59'38.223"N
MU	43	Sumter	43002000	971775	Minor	0.082129964	0.044520548	0.02128	0.000078	80°20'40.91"W 33°56'12.874"N
MU	43	Sumter	43000800	1112854	Major	0.082129964	0.037100949	0.13333	0.000406	80°21'13.686"W 33°56'48.341"N
MU	43	Sumter	43001802	1112939	Major	0.082129964	0.17820324	0.07692	0.001126	80°28'8.022"W 33°57'40.873"N
MU	43	Sumter	43000202	1113052	Major	0.082129964	0.134883721	0.25	0.002769	80°28'24.357"W 34°2'19.433"N
MU	43	Sumter	43000202	1114998	Major	0.082129964	0.134883721	0.25	0.002769	80°28'5.612"W 34°1'45.121"N
MU	43	Sumter	43002000	1115012	Major	0.082129964	0.044520548	0.4	0.001463	80°20'19.862"W 33°55'38.331"N
MU	43	Sumter	43002000	1115018	Major	0.082129964	0.044520548	0.4	0.001463	80°20'21.455"W 33°55'45.618"N
MU	43	Sumter	43001802	1115488	Major	0.082129964	0.17820324	0.07692	0.001126	80°30'18.19"W 33°56'53.182"N
MU	43	Sumter	43000800	1115493	Major	0.082129964	0.037100949	0.13333	0.000406	80°20'43.598"W 33°56'46.313"N

Table 5.15 Williamsburg Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
LR	45	Williamsburg	45970100	663992	Minor	0.125615764	0.343283582	0.04762	0.002053	79°48'18.446"W 33°42'44.949"N
LR	45	Williamsburg	45970600	664214	Minor	0.125615764	0.603603604	0.01538	0.001166	79°52'16.676"W 33°43'13.806"N
LR	45	Williamsburg	45970200	664699	Minor	0.125615764	0.333333333	0.125	0.005234	79°31'37.548"W 33°45'7.523"N
LR	45	Williamsburg	45970700	664707	Minor	0.125615764	0.52688172	0.07143	0.004727	79°59'8.802"W 33°34'51.861"N
LR	45	Williamsburg	45970200	886292	Major	0.125615764	0.333333333	1	0.041872	79°35'46.479"W 33°44'32.482"N
LR	45	Williamsburg	45970600	886298	Major	0.125615764	0.603603604	1	0.075822	79°50'3.847"W 33°40'13.431"N
LR	45	Williamsburg	45970100	886351	Major	0.125615764	0.343283582	1	0.043122	79°48'28.074"W 33°43'25.95"N
LR	45	Williamsburg	45970700	886399	Major	0.125615764	0.52688172	1	0.066185	79°59'55.274"W 33°34'57.375"N
LR	45	Williamsburg	45970700	886401	Major	0.125615764	0.52688172	1	0.066185	80°1'11.387"W 33°35'55.892"N
LR	45	Williamsburg	45970200	886447	Major	0.125615764	0.333333333	1	0.041872	79°34'39.849"W 33°45'32.241"N
LR	45	Williamsburg	45970600	886449	Major	0.125615764	0.603603604	1	0.075822	79°54'16.283"W 33°46'47.508"N
LR	45	Williamsburg	45970100	887561	Major	0.125615764	0.343283582	1	0.043122	79°48'28.584"W 33°43'19.315"N

Table 5.16 York Sampled Road Reaches

Strata	Code	County	Tractc	RoadSamp	RoadType	CountyProb	TractProb	RoadProb	Prob	GPS Coordinates
UU	46	York	46061501	960538	Minor	0.179614486	0.048154093	0.0073	0.000063	81°13'4.431"W 35°3'34.108"N
UU	46	York	46060401	961759	Minor	0.179614486	0.026272578	0.01333	0.000063	81°2'8.897"W 34°54'34.161"N
UU	46	York	46060200	962085	Minor	0.179614486	0.029142109	0.0125	0.000065	81°0'38.522"W 34°55'28.14"N
UU	46	York	46060901	964798	Minor	0.179614486	0.064078283	0.00629	0.000072	80°58'21.692"W 34°58'40.029"N
UU	46	York	46061501	1106742	Major	0.179614486	0.048154093	0.15385	0.001331	81°15'24.265"W 35°2'19.737"N
UU	46	York	46060401	1107242	Major	0.179614486	0.026272578	0.4	0.001888	81°1'55.266"W 34°55'27.079"N
UU	46	York	46060200	1107247	Major	0.179614486	0.029142109	0.22222	0.001163	80°59'50.948"W 34°56'24.268"N
UU	46	York	46060200	1107653	Major	0.179614486	0.029142109	0.22222	0.001163	80°59'49.047"W 34°56'17.758"N
UU	46	York	46060901	1107732	Major	0.179614486	0.064078283	0.04545	0.000523	80°58'21.448"W 34°56'30.651"N
UU	46	York	46061501	1109019	Major	0.179614486	0.048154093	0.15385	0.001331	81°15'24.168"W 35°2'18.499"N
UU	46	York	46060401	1109162	Major	0.179614486	0.026272578	0.4	0.001888	81°1'52.415"W 34°55'26.364"N
UU	46	York	46060901	1109176	Major	0.179614486	0.064078283	0.04545	0.000523	80°59'9.939"W 34°58'32.403"N

Table 5.17 April Small Sample

Strata	Code	County
UU	4	Anderson
UR	37	Oconee
MU	32	Lexington
MR	38	Orangeburg
LU	10	Charleston
LR	15	Colleton

CHAPTER 6

CONCLUSION

Now that the sample has been produced it will remain in use for the next five years until it is resampled in 2023 (barring any changes to the historical application of the survey). First, the survey for April 2018 will be conducted with the smaller sample of six. Then in June 2018 the entire sample of sixteen will be conducted. Data will be analyzed after this survey, and after each subsequent year's two samples (Grego 4).

In the future, sampling could be simplified with some minor changes. Most importantly the variable ObjectId should be used as the identification code variable instead of MSLINK and ID. This would prevent the issue that arose of mismatched data due to identification variables overlapping in the Statewide Highway and Statewide Other Roads datasets. Secondly, addressing the issue of numeric variables being read in inconsistently as both numeric and character should be done prior to any analysis. It is easy to lose track of records with a dataset of this magnitude in SAS without immediately noticing. Further, it is not suggested to match on numeric variables in SAS. This presents a problem if County is numeric as it is necessary to simplify multiple datasets down to only those that include the sampled counties, for instance removing counties with less than the NHTSA's threshold for traffic fatalities .

The analysis may also evolve as the nation increases in average usage of safety belts. Investigations into some of the road types purposely left out of this study may be

more warranted, as South Carolina nears the nationwide 89.7% usage (Li, and Pickrell 1). Comparisons between roads with low speed limits, like the dirt roads and mall parking lots that were eliminated from this study, and roads with higher speed limits may be more important as awareness increases. The focus could also shift to motorcycles, which while currently studied were not included until 2006 (Grego 7). As less motorcycles are observed, statistical inference on this population is more difficult. However, if the interest shifted to motorcycles, a change in sampling could give higher counts completed to get higher total counts for motorcycles. Locations could be sampled with weights directly related to motorcycle usage and counters could remain in location for longer durations. This would result in a need to reduce locations to offset costs. Nevertheless, as the safety belt usage rate nears 100%, the focus of the study is likely to change directions and the sampling will need to adjust to suit the new aims.

REFERENCES

"South Carolina Department Of Public Safety". Buckleupsc.Com, 2013, http://www.buckleupsc.com/seat_belt_law.asp. Accessed 13 Mar 2018.

Grego, John. South Carolina Statewide Survey Of Safety Belt Use: June 2008. Department Of Statistics At University Of South Carolina, Columbia, SC, 2008, pp. 3-20, http://www.scdps.gov/oea/attachments/SafetyBelt_Report_2008.pdf. Accessed 13 Mar 2018.

Li, R., and T.M. Pickrell. Seat Belt Use In 2017—Overall Results. National Highway Traffic Safety Administration, Washington, DC, 2017, pp. 1-4, https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812465. Accessed 13 Mar 2018.

Moore, David S., and William I. Notz. Statistics: Concepts and Controversies. 8th ed., W.H. Freeman and Company, 2012.

“Spatial Join.” ArcGIS for Desktop, ArcMap, desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/spatial-join.htm.

United States, Congress, South Carolina Seat Belt Law. South Carolina Law “Section 56-5-6520.”, SC Code, 2012.

APPENDIX A

R CODE: FUNCTIONS FOR SAMPLING COUNTIES

ADAPTED FROM STATLAB 2003

```
funct.2<-function(Nh,p){  ## for sampling two counties from strata
n<-Nh
pii<-rep(0,n)  ## create and assign weights
for(i in 1:n){
sumj<-0
for(j in 1:n){
if(j!=i){
sumj<-sumj+p[j]/(1-p[j])
}
}
pii[i]<-p[i]*(1+sumj)
}
pii
}

funct.3<-function(Nh,p){  ## for sampling three counties from strata
n<-Nh
pii<-rep(0,n)  ## create and assign weights
```

```
for(i in 1:n){
sumj<-0
for(j in 1:n){
if(j!=i){
sumk<-0
for(k in 1:n){
if((k!=i)&(k!=j)){
sumk<-sumk+p[k]/(1-p[j]-p[k])
}}
sumj<-sumj+p[j]*(1+sumk)/(1-p[j])
}}
pii[i]<-p[i]*(1+sumj)
}
pii
}

funct.4<-function(Nh,p){  ## for sampling four counties from a strata
n<-Nh
pii<-rep(0,n)  ## create and assign weights
for(i in 1:n){
sumj<-0
for(j in 1:n){
if(j!=i){
```

```
sumk<-0
for(k in 1:n){
if((k!=i)&(k!=j)){
suml<-0
for(l in 1:n){
if((l!=i)&(l!=j)&(l!=k)){
suml<-suml+p[l]/(1-p[j]-p[k]-p[l])
}}
sumk<-sumk+p[k]*(1+suml)/(1-p[j]-p[k])
}}
sumj<-sumj+p[j]*(1+sumk)/(1-p[j])
}}
pii[i]<-p[i]*(1+sumj)
}
pii

## creating sample using weights
Sprob1=funct.2(length(strata.name[strata.name=="LR"]),
(Csamp[(strata.name=="LR"),4]))
Sprob2=funct.3(length(strata.name[strata.name=="LU"]),
(Csamp[(strata.name=="LU"),4]))
Sprob3=funct.2(length(strata.name[strata.name=="MR"]),
(Csamp[(strata.name=="MR"),4]))
```

```
Sprob4=funct.3(length(strata.name[strata.name=="MU"]),
(Csamp[(strata.name=="MU"),4]))
Sprob5=funct.2(length(strata.name[strata.name=="UR"]),
(Csamp[(strata.name=="UR"),4]))
Sprob6=funct.4(length(strata.name[strata.name=="UU"]),
(Csamp[(strata.name=="UU"),4]))
```

APPENDIX B

R CODE: FUNCTION FOR SAMPLING TRACTS

ADAPTED FROM STATLAB 2003

```
funct.4<-function(Nh,p){  ## for sampling four census tracts
n<-Nh
pii<-rep(0,n)  ## creating and assigning weights
for(i in 1:n){
sumj<-0
for(j in 1:n){
if(j!=i){
sumk<-0
for(k in 1:n){
if((k!=i)&(k!=j)){
suml<-0
for(l in 1:n){
if((l!=i)&(l!=j)&(l!=k)){
suml<-suml+p[l]/(1-p[j]-p[k]-p[l])
}}
sumk<-sumk+p[k]*(1+suml)/(1-p[j]-p[k])
}}
sumj<-sumj+p[j]*(1+sumk)/(1-p[j])
```

```
}}
pii[i]<-p[i]*(1+sumj)
}
pii
}

## creating sample using weights
Tsamp=data.frame(samp.name,samp.code,unlist(sample.CTweights),samp.wtract)
Tprob=by(Tsamp, Tsamp[,2], function(Tsamp) funct.4(4, (Tsamp[,4])))
```

APPENDIX C

R CODE: FUNCTION FOR SAMPLING ROAD REACHES

ADAPTED FROM STATLAB 2003

```
samp3=function(x){
smaj=x[,8][(x[,7]=="Major")]  ## creating sampling pool of major roads
smin=x[,8][(x[,7]=="Minor")]  ## creating sampling pool of minor roads
sampf=c(sample(smaj,2),sample(smin,1))  ## SRSs
return(sampf)
}

## creating sample
SampleRoad=by(TractRoad,TractRoad[,1],samp3)
```

APPENDIX D

R CODE: SMALL SAMPLING BY STRATA

```
## creating pools for strata

UU = c("4:Anderson","23:Greenville","42:Spartanburg","46:York")

UR = c("11:Cherokee","37:Oconee")

MU = c("2:Aiken","32:Lexington","43:Sumter")

MR = c("16:Darlington","38:Orangeburg")

LU = c("10:Charleston","18:Dorchester","26:Horry")

LR = c("15:Colleton","45:Williamsburg")


## sampling each strata

UU_select = sample(UU,1)

UR_select = sample(UR,1)

MU_select = sample(MU,1)

MR_select = sample(MR,1)

LU_select = sample(LU,1)

LR_select = sample(LR,1)
```

APPENDIX E

SAS CODE

```
*Import Primary Road file;

libname Frame17 "\\uscmars\statlab$\Projects\Paid\Safety Belt\Sampling 2017"; run;


PROC IMPORT OUT= WORK.statewide

        DATAFILE= "\\uscmars\statlab$\Projects\Paid\Safety Belt\Sampling

2017\Statewide_Highways\Statewide Highways Join\CB_and_SWH_join.dbf"

        DBMS=DBF REPLACE;

    GETDELETED=NO; RUN;


*Import Secondary Road file;

PROC IMPORT OUT= WORK.Local

        DATAFILE= "\\uscmars\statlab$\Projects\Paid\Safety Belt\Sampling

2017\Statewide_Other_Roads\Statewide Other Join\CB_and_SWO_join.dbf"

        DBMS=DBF REPLACE;

    GETDELETED=NO; RUN;


data statewide1; format county_id f2.0; set statewide;

*Use same ID variable for both data sets;

objectid=MSLINK+250000;
```

```
keep county_id route_type route_numb route_dir beg_milepo end_milepo route_id feature_ty
    objectid street_nam tractce;
*Remove counties with few traffic fatalities;
if county_id not in (.,0,1,3,5,6,9,13,19,20,24,25,31,33,34,35,36,41,44); run;

*For some reason, county_id=1 is kept in the database;
proc freq data=statewide1;
table county_id; run;

data local1;format county_id f2.0;  set local;
objected=id;
keep county_id route_type route_numb route_dir beg_milepo end_milepo route_id feature_ty
    objectid street_nam tractce;
*Remove counties with few traffic fatalities and not in SMSA;
if county_id not in (.,0,1,3,5,6,9,13,19,20,24,25,31,33,34,35,36,41,44); run;

*For some reason, county_id=1 is kept in the database here too;
proc freq data=local1;
table county_id; run;

*Combine counties, get rid of county_id 1 (Abbeville);
data all; set statewide1 local1; if county_id gt 1;
*Create 8-digit county/tract character variable with leading zero;
county_id_c=put(county_id,z2.);
```

```
new_tractce=county_id_c||tractce; run;


data ccode29;
input county_ID strat $ pop vmt @@;
*Set up match merge with WORK.ALL on character variable rather than numeric variable;
county_id_c=put(county_id,z2.);
datalines;
2 MU 160.1 4.95
4 UU 187.1 5.69
7 LU 162.2 3.46
…
46 UU 226.1 6.15; run;


proc sort data=ccode29; by strat; run;
proc means data=ccode29 noprint; var vmt; by strat; output out=outa sum=sumvmt; run;


/* Clean up output data set       */
data outb; set outa;
drop _type_ _freq_;
run;
proc print data=outb; run;


proc sql;
select strat,16*sumvmt/sum(sumvmt) as prop from outb;  quit;
```

```
/* One more sort just to be safe     */
proc sort data=outb; by strat; run;

*Create CountyWeight.txt from WORK.C to select 16 counties using ....R;
data c; merge ccode29 outb; by strat;  drop pop;
vmtprop=vmt/sumvmt; run;

*Add strata information (and population and VMT) in WORK.ccode29 to road reach data set
work.all;
proc sort data=ccode29; by county_id_c; run;
proc sort data=all; by county_id_c; run;
data all1; merge all ccode29; by county_id_c; run;

*Check what is in feature-ty and route-ty;
proc freq data=all;
          tables route_type * feature_ty / NOPERCENT NOROW NOCOL; run;

*Clean up data set;
data all2; set all1;
if compress(route_type) not in ("S-");
if route_type not in ("UD","");
if tractce ne "";
if compress(feature_ty) not in ("City_Unpaved_Sec","City_Unpaved_CoRd","PRIVATE","Private",
```

```
"S_Ramp","UD_Unpaved","UD_Unpaved_City","Undocumented","Unpaved_CoRd","Unpaved_S
ec","Pri_Div",
"Pri_Multi","Pri_TwoLane");
reach_length=end_milepo-beg_milepo; run;

*This version has 29 counties;
data Frame17.sampleframe; set all2;
if route_type="L-" then roadtype="Minor";
else roadtype="Major";
county=county_id;
keep county roadtype tractce objectid new_tractce county_id_c; run;

/* reducing sampling frame to 16 counties */
*Maybe only need badcty/goodcty is ctidfp00 is character and cty is coded 1,3,5,7...;
data Frame17.sample16;
set Frame17.sampleframe;
if county_id_c in ('15','45','10','18','26','38','16','43','32','02','11','37','42','46','23','04'); run;

*Check that all 16 counties are captured;
proc freq data=Frame17.sample16;
table county_id_c; run;

/*Fix the problem tracts with fewer than 2 intersection samples*/
proc import out=tract16_2
```

```
        DATAFILE= "\\uscmars\statlab$\Projects\Paid\Safety Belt\Sampling
2017\2016_Statewide_Traffic_Points\CB_and_TP_join.dbf"
        DBMS=DBF REPLACE;
    GETDELETED=NO; RUN;


data tract16_3; set tract16_2;
            if tractce = '001203' then tractce = '001805';
            if tractce = '001205' then tractce = '001204';
            if tractce = '002007' then tractce = '002006';
…
            if tractce = '070702' then tractce = '070701';
   county_c=put(input(county,best2.),z2.);
            new_tractce=county_c||tractce;
*Numeric version of county not created, but can be created in future;
            keep county_c id1 tractce new_tractce latitude longitude; run;




data tract16_4; set tract16_3;
            if new_tractce = '02020801' then new_tractce = '02020802';
            if new_tractce = '02022001' or new_tractce = '02022002' or new_tractce = '02980100'
                    then new_tractce = '02022100';
            if new_tractce = '04000300' then new_tractce = '04000200';
```

```
        …
        if new_tractce = '46061202' then new_tractce = '46061203';
        keep county_c id1 tractce new_tractce latitude longitude; run;

proc sort data=tract16_4; *by county tractce; by new_tractce; run;

proc sort data=frame17.sample16; *by county tractce; by new_tractce; run;

data tract16_5;
merge frame17.sample16 (IN=tract16) tract16_4 (IN=tractaadt);
*by county tractce; by new_tractce;
if tract16=1 and tractaadt=1;
*cTract=input(new_ctract, best32.); run;

*Check on number of counties and tracts;
proc sql;
select count(distinct county_c)as ctycount, count(distinct new_tractce) as tractcount from
tract16_5; quit;

*Identify tracts with their counts;
proc sql;
create table tract16_count as select count(new_tractce) as countce, county, new_tractce,
roadtype, ID, ID1 from tract16_5 group by new_tractce; quit;
```

```
*checking counts of major and minor in each tract--have some with less than 2 major
proc freq data=tract16_count;
table new_tractce*roadtype; run;

*Census tracts are sampled proportional to AADT;
*Set up for sampling tracts.  Okay since county and tractce are both character;
*Mean AADT within each tract;
proc sql;
create table meanaadt as select county, new_tractce, mean(countce) as maadt from
tract16_count
group by county, new_tractce; quit;

*Proportional AADT within each tract;
proc sql;
create table propaadt as select county, new_tractce, maadt, maadt/sum(maadt) as wtract from
meanaadt
group by county; quit;

*propaadt was exported to file TractWeights.txt in "R folder for use by R program SampeTracts
to sample census tracts and save them in ...;

*road 16 is used in R as Road sampling pool;
data road16;
set tract16_count;
```

```
keep id roadtype new_tractce; run;


*Roadsamp.txt is the set of 192 sampled roads from R;
PROC IMPORT OUT= WORK.RSAMP
        DATAFILE= "Z:\Projects\Paid\Safety Belt\Sampling 2017\RoadSamp1.txt"
        DBMS=TAB REPLACE;
    GETNAMES=YES;
    DATAROW=2;  RUN;
* Changing tract to character;
data Rsampc; set Rsamp;
           Tractc=put(input(Tract,best32.),z8.);
           keep RoadSamp RoadType Tractc; run;


proc sql;
create table RoadCount as select id, new_tractce, roadtype, count(id) as Rcount from Road16
group by new_tractce, roadtype; quit;


*Compute roadweights;
data roadweight;
set RoadCount;
if roadtype="Major" then roadprob=2/Rcount;
else roadprob=1/Rcount; run;


*Add roadweights to road sample;
```

```
proc sql;
create table Rsampweight as select * from Rsampc left join Roadweight
on (Tractc=new_tractce) and (RoadSamp=id); quit;


data Rsampweight2;
set Rsampweight;
keep Tractc RoadSamp roadtype roadprob;
run;
PROC IMPORT OUT= WORK.Tractprob
        DATAFILE= "Z:\Projects\Paid\Safety Belt\Sampling 2017\R\TractSampleProb.txt"
        DBMS=TAB REPLACE;
    GETNAMES=YES;
    DATAROW=2;  RUN;


*Changing tract to character;
data Tractprob2;
set Tractprob;
Tract2=put(input(Tract,best32.),z8.); run;


*Add Tract weights to road weights;
proc sql;
create table SampProb as select * from Rsampweight2 as a right join Tractprob2
on a.Tractc=Tract2; quit;
```

```
*fixing County Names;
data SampProb2; set SampProb;
        drop County; run;

data SampProb2; set SampProb2;
        if Code = '45' then County = 'Williamsburg';
        if Code = '2' then County = 'Aiken';
        if Code = '4' then County = 'Anderson';
        …
        if Code = '46' then County = 'York'; run;

proc sql;
create table SampProb3 as select Strata, Code, County, Tractc, RoadSamp, Roadtype,
CountyProb, TractProb, RoadProb,
(CountyProb*TractProb*roadprob)as Prob from SampProb2; run;

proc sql;
select sum(distinct countyprob)from sampprob3;
run;
ods rtf file="Z:\Projects\Paid\Safety Belt\Sampling 2017\sampprob.rtf";
proc print data=sampprob3; run;
ods rtf close;

/* Create more comprehensive output listing including street names to assist  */
```

```
/* with marking up maps */;
libname  samp2017 "z:\Projects\Paid\Safety Belt\Sampling 2017"; run;
PROC IMPORT OUT= WORK.samplong
        DATAFILE= "Z:\Projects\Paid\Safety Belt\Spring 2013\sampprob.txt"
        DBMS=TAB REPLACE;
    DELIMITER='09'x;
    GETNAMES=YES;
    DATAROW=2;  RUN;
*/
/* Re-ran initial creation of data set all.  Let's merge it and save road names */
/* Add latitude and longitude as well */
data samplong;
Merge Sampprob3(in=sp3) Table2(in=all);
if sp3;
by Tract;

/* Some have multiple lat/longs, keeping only the last one*/
proc sql;
create table samplong as select a.*, b.latitude, b.longitude from Sampprob3 as a left join
Tract16_5 as b on a.RoadSamp=b.objectID; quit;

proc sort data=samplong out=samplong2;
          by Tractc RoadSamp; run;
```

```
data samplong2; set samplong2;
        by Tractc RoadSamp ;
        if last.RoadSamp; run;

/* Adding Street Name and Route ID */
proc sql;
create table road_names as select a.*, b.route_id, b.street_nam, from samplong2 as a left join
all as b
on a.roadsamp=b.objectid order by a.county, a.roadsamp; quit;

proc sql;
create table road_names as select a.strata, a.Code, a.County, a.Tractc, a.RoadSamp, a.Roadtype,
a.CountyProb, a.TractProb, a.RoadProb, a.prob, a.latitude, a.longitude, b.route_id, b.street_nam
from samplong2 as a left join all as b
on a.roadsamp=b.objectid order by a.county, a.roadsamp; quit;

/*Export it*/
ods rtf file="Z:\Projects\Paid\Safety Belt\Sampling 2017\sampprobRoadNames.rtf";
proc print data=road_names; run;
ods rtf close;
```

APPENDIX F

EXAMPLE COUNTERS' MAP