What proportion of the U.S. is within a mile of a road?

Nicholas J. Horton

Department of Mathematics and Statistics
Smith College
Northampton, MA

March 22, 2011

nhorton@smith.edu
http://www.math.smith.edu/∼nhorton

Introduction

goal	motivate the estimation of unknown population parameters
activity	estimate the proportion of the continental United States that is within a mile of a road
process	repeatedly sample latitudes and longitudes and explore properties of that location using an internet mapping service
outcome	individual as well as class-wide estimates (and associated confidence intervals) for the proportion
technology	straightforward to implement in general purpose statistical software

Acknowledgements

- derived from an activity developed by Danny Kaplan (Macalester College)
- work supported by the MOSAIC project (NSF funded community of educators working to develop a new way to introduce mathematics, statistics, computation and modeling to students in colleges and universities, www.mosaic-web.org
- more information and downloadable resources (handouts, sample data, R code and these slides) available from www.math.smith.edu/~nhorton/roadless

Background on the activity

- population parameter: proportion of continental US within 1 mile of a road
- relevant for conservation and land use planning
- hard to measure in population
- sampling is an attractive approach
- technology facilitates manual "data scraping" on the web

USGS project

USGS
science for a changing world

Distance to Nearest Road in the Conterminous United States

A New Dataset

The USGS Geographic Analysis and Monitoring (GAM) program has developed a national, high resolution dataset that gives the distance to the nearest road every 30 meters across the conterminous 48 states. This work provides the first unified national picture of roadless space, vehicular accessibility, and intensity of road construction.

The new dataset is the first member of the National Overview Road Metrics (NORM) family of road related indicators. This indicator measures straight-line or Euclidean distance (ED) to the nearest road, and is given the compound name *NORM ED*. NORM ED data can be viewed and downloaded from the transportation section of the web viewer for *The National Map*, http://nationalmap.usgs.gov. The fullresolution dataset for the conterminous states is made up of 8.7 billion values.

The North American Road Network

Overview

The road network of the United States is one of the largest human constructions on Earth. It consists of more than 4 million miles of mapped roads, plus many additional utility roads, 4-wheel-drive trails, and private routes. The documented roads and their rights-of-way occupy approximately 1% of the land area of the United States, roughly the area of South Carolina.

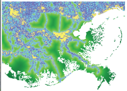
Roads are the circulatory system of our culture. They are used to bring raw materials to processing sites, to distribute processed goods, and to carry people to their schools, workplaces, jobs, and friends. The value of roads is reflected in the significant size of the economic sectors—energy, automotive manufacturing, mining, and construction, among others—that support their building and use.

The U.S. road network is little studied as an integrated object. Spatial relationships between the network and intervening roadless areas are important to ecological and hydrological resources. The NORM indicators provide basic descriptions of the association of roads with their surroundings. NORM ED, in particular, focuses not on the characteristics of the road network itself, but rather on the extent of spaces between roads.

The National Distance to Road Image

Properties of the Image

The background on this page is an image of average values of distance to road (DTR) for cells measuring 1,020 m (0.63 miles, nominally 1 km) on each side; 7.5 million 1-km cells are displayed. Each cell averages 1,156 DTR values from the full-resolution 30-m dataset. The same image is reproduced in more vivid color on page 2.



NORM-ED 1-km resolution image of southern Louisiana. Baton Rouge is the yellow patch in the northwest corner and New Orleans the yellow patch in the center. Lakes Pontchartrain and Maurepas form the large green patch north of New Orleans, interrupted by only two roads. The Mississippi Delta extends to the southeast. The barrier islands stretching northeast and southwest from the delta are some of the places most remote from roads in the conterminous United States. Other green areas are nearly roadless coastal swamps.

Patterns of Open Space and Road Density

Areas of lowest distance to road (highest road density) appear in yellow. Most low DTR areas are in cities, but there are exceptions. For example, energy extraction activities have developed dense road networks in the oil and gas fields of West Texas and the Texas Panhandle; these appear as yellow arcs. The northern of two adjacent bright yellow areas in north central New Mexico is the product of road construction in advance of settlement in the city of Rio Rancho.

Moderately low DTR (moderately high road density) is colored blue. These conditions typically occur in areas where terrain and vegetation are not impediments to road building, and where there is motivation—such as agriculture—for making land highly accessible. Blue areas occur in all states, and are particularly prominent across the Great Plains; along the Snake River Plain in Idaho; in the Central and Imperial Valleys of California; and to the west of the Cascade Mountains in the Pacific Northwest. The Atlantic Seaboard, outside of its many cities, falls in this DTR range.

High DTR (low road density) is represented in shades of blue-green or turquoise. The alternating valleys and ridges of the Appalachian Mountains create a pattern of blue (valleys) and turquoise (ridges). A broad swath of blue-green covers much of the arid lands of the West, from the sand hills of Nebraska to the deserts of Arizona, New Mexico, Nevada, and southeastern California. Here agricultural productivity is low, and economic motivation for building a dense road network is generally lacking. Across the Interior West, where higher road densities (blue or yellow) occur, there is an association with water availability for agriculture (along the Rio Grande in New Mexico, for example), natural resource extraction, or dense population.

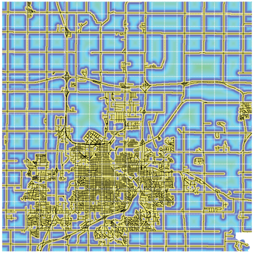
Shades of green represent areas remote from roads. Most of these occur because of severe accessibility issues. Steep slopes (Rocky Mountains, Sierra Nevada), swamps (southern Florida and Louisiana), and in some places climate (northern Maine and Minnesota), have made road building difficult and expensive. Many of these places are preserved as National Parks (Yellowstone, Everglades) and Wilderness Areas. Nationally, these roadless areas are scarce, and they occur preferentially in the West.

Applications

Environmental Assessment

Roads and traffic affect natural resources in dozens of ways. Among these are elimination of forest canopy, elevation of temperature, introduction of vehicular noise and pollution, diversion and concentration of natural drainage, production of dust, introduction of invasive species, and collision of vehicles with animals large and small. Roads also serve as the primary mechanism for conveyance of humans into the landscape, with consequent resource extraction, initiation and extinction of fires, construction of dwellings, introduction of domestic animals, and building of additional roads. It has been estimated that roads ecologically affect 22% or more of the land area of the conterminous United States. Because the effects of roads extend beyond their rights-of-way, a distance-to-road measure is particularly helpful in estimating the lateral reach and areal extent of ecological and hydrological effects of roads.

Imposed fragmentation of the landscape is generally detrimental to ecological integrity. NORM ED measures human-induced fragmentation by roads and ignores other fragmentation types, including natural fragmentation. Because roads are the most common source of human-induced fragmentation, NORM ED provides a resource for studies of human impacts on a national scale and a basis for comparing landscape patterns manipulated by humans to natural patterns.



Full-resolution NORM ED image of Sioux Falls, South Dakota. Light lines are boundaries of 1,020-m cells, the unit of averaging for the national image. The 1-mile section-line road grid is visible outside the city. Interstate 90 runs east-west just above image center.

Land Use and Land Cover Research

In our society, human presence and vehicular access by road are tightly coupled. New roads often are built to support new land uses. As the road network changes, so does the mosaic of intervening roadless spaces. Scientific study of the processes of road-network extension and densification, and their coupling to demographic and economic conditions, is just beginning. Their study may provide valuable insights into environmental challenges that may lie ahead.

Source Data and its Limitations

The NORM ED dataset is derived from the Bureau of Transportation Statistics/ Geographic Data Technology (BTS/GDT) roads dataset, which in turn was derived from the Bureau of Census TIGER files for the 2000 census. The origins of the data range from pre 1990 USGS quadrangle maps at 1:100,000-scale to recent datasets built using highly accurate Global Positioning System technology. Because the data have such mixed origins, they cannot be interpreted as uniform in resolution, quality, or age. Nevertheless, this is the most current dataset publicly available that describes the road network of the United States.

Values Over Water

The downloadable NORM ED dataset includes DTR values over lakes and offshore. Over the oceans and Great Lakes, DTR was calculated to a range of 60 km, with extended ranges where necessary to cover roadless U.S. islands. The centerfold map does not depict DTR values for water areas that are outside county boundaries; among these white areas are oceans, the Great Lakes, and various estuaries. Other water areas, such as Utah's Great Salt Lake, fall within county boundaries and are depicted in DTR colors.

More Information

For more information on NORM ED, contact:

Raymond D. Watts
USGS Fort Collins Science Center
2150 Centre Ave Bldg C
Fort Collins, CO 80526-8118
e-mail: rwatts@usgs.gov
phone: 970-226-9378

Organization

- students given handout (included on website) describing process
- work in groups of 2 (a la "pair programming") to proofread and check results
- instructor supplied R-code used to generate spreadsheet containing random latitudes and longitudes and framework for data entry
- students open two windows: one web browser and one for Excel
- variables to code: `incontinent` (0 or 1), `within1mile` (0 or 1), and `location` (character)

R code

```
nsamp = 50
long = -runif(nsamp, min=65, max=130)
lat = runif(nsamp,min=25,max=50)
data = cbind(sample=1:nsamp,
   latitude=round(lat, 4),
   longitude=round(long, 4),
   incontinent=c(rep(NA, nsamp)),
   within1mile=c(rep(NA, nsamp)),
   location=character(nsamp),
   notes=character(nsamp))
write.csv(data,file="roadless.csv")
```

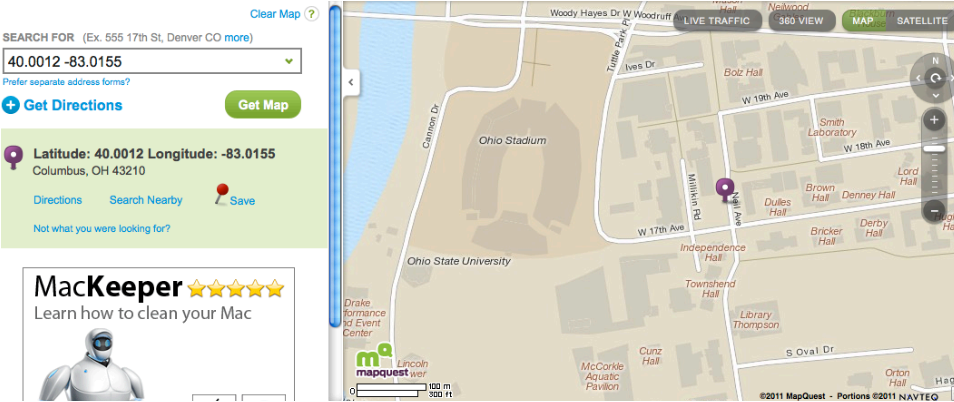
Mapquest.com

- search window allows user to jump to a particular point by providing latitude and longitude
- can zoom in and out as appropriate to find where you are
- legend on bottom left provides scale (1 mile = 5280 feet)
- R used to generate spreadsheet containing random latitudes and longitudes and framework for data entry in Excel using code provided by instructor

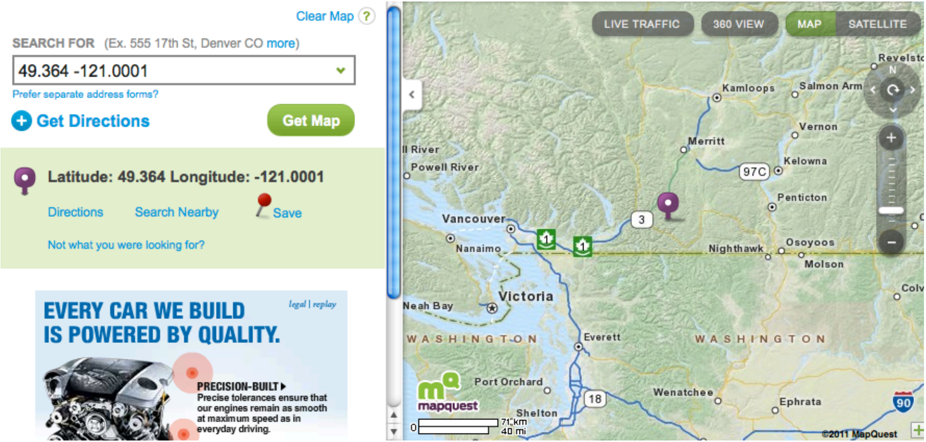
Example

- Imagine that our first random location was `40.0012 -83.0155`
- Is this within the continental US? If so, is it within a mile of a road?
- Let's cut and paste these values from our spreadsheet into Mapquest.

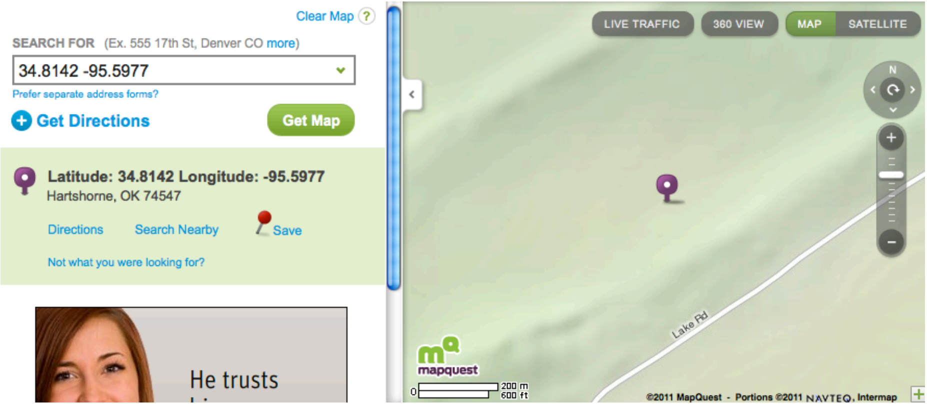
Example 1: the Ohio State University Statistics Dept



Example 2: British Columbia



Example 3: Oklahoma



Analysis and deliverables

- students collect data from 50 samples
- students analyze results in R, calculating their group's interval estimate
- results get pooled for the class as a whole, yielding a smaller interval
- reinforces many important questions about design and estimation

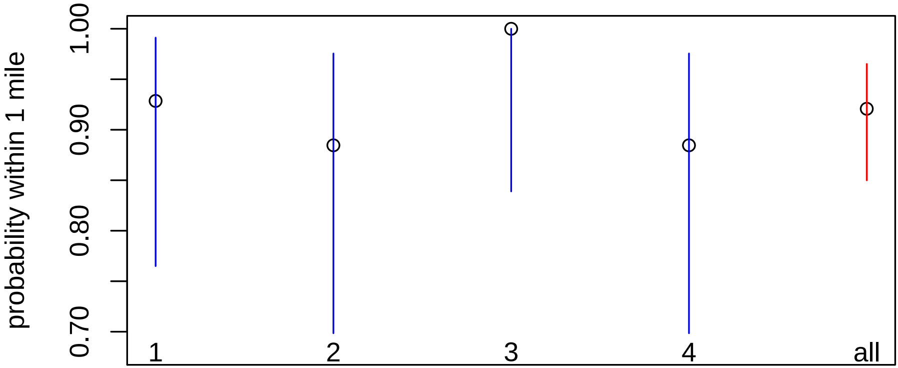
Analyses of a sample of size 50

```
> ds = read.csv("roadless.csv", stringsAsFactors=FALSE)
> table(ds$incontinent)
 0  1
22 28

> table(ds$location[ds$incontinent==0])

Atlantic          Canada  Gulf of Mexico          Mexico
      10               1               2               2
 Pacific           water
       6               1
```

Analyses of a sample of size 50

```
> smallds = subset(ds, incontinent==1); rm(ds)
> table(smallds$within1mile)
 0  1
 2 26
> binom.test(26, 28)
Exact binomial test
data:  26 and 28
number of successes = 26, number of trials = 28, p-value =
alternative hypothesis: true probability of success is not
95 percent confidence interval:
 0.7649652 0.9912295
sample estimates:
probability of success
             0.9285714
```

Results from 4 groups (plus combined)



Discussion

- straightforward approach to estimation of an unknown population parameter
- hard to measure the true value
- the parameter is moderately interesting
- estimator is creative
- easy to implement in a single class (if taught in computer classroom)
- helps reinforce the use of technology in the course
- brings in many design questions (how do I measure a mile? Should unlabeled paths be included? What if I land in the middle of a 3 mile wide lake?)

Extensions

- can be done outside of class (if lab space not available)
- can audit student work to ensure that they are appropriately coding their locations
- students asked to summarize and assess assumptions used
- note that the world is not flat (duh!): nice connections to trig in terms of improving the sampling (see the `rgeo()` function within the `mosaic` package
- other ideas?

What proportion of the U.S. is within a mile of a road?

Nicholas J. Horton

Department of Mathematics and Statistics
Smith College
Northampton, MA

March 22, 2011

nhorton@smith.edu
http://www.math.smith.edu/∼nhorton