4 Lessons in Creating a New Trade Area Methodology for Retail Stores

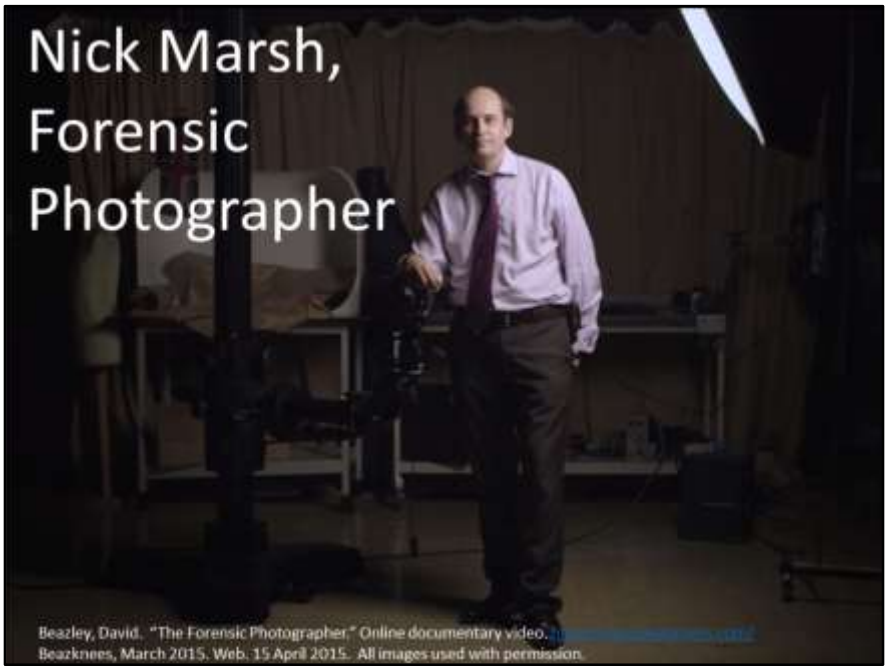
Susan Zwillinger, GISP
Principal Consultant

July 23, 2015

4CGeoWorks

Abstract

- Four lessons from the field of forensics will summarize tips to create trade areas for a variety of retail store types.
- **The problem:** Simple trade area methods don't work well when you have thousands of convenience stores to study.
- **The solution:** Using Circle K and other c-stores as an example, this presentation will illustrate a new methodology for creating trade areas using Esri Business Analyst.
- The presentation will discuss
  - the importance of trade area definition and classification,
  - using the right tools,
  - building a methodology and
  - finding the hidden knowledge from analytics.



Meet Nick Marsh, a forensic photographer in London. I was introduced to Nick through a video created by David Beazley. The point of the video was to highlight that the greatest tool for a forensic photographer is his enquiring mindset. In essence, it is the quality of curiosity, one that is eager to seek information and acquire knowledge, that can make the difference between a good forensic photographer and a great one. The same is true for GIS.

http://aeon.co/video/society/the-forensic-photographer-why-expertise-still-matters-aeon/

I think there is a lot that you can learn about GIS analysis by understanding how it is similar to forensics. Forensics is the application of scientific methods and techniques to the investigation of crime. It involves the process of gathering and explaining information about the past which is then used in a court of law. Forensic scientists must determine what is **relevant**, what is and is not in a crime scene, and then they must conduct tests and analyze the data in order to **interpret the results** and **communicate** this information in a way that the average person can understand the information. This process is very similar to what a GIS analyst does for market analysis.

In this presentation, I would like to teach you four principles from Nick's profession that translate to GIS analysis and, more specifically, to a process that I developed to create a new trade area methodology for convenience based retail stores. These techniques can be modified to suit other retailers as well.

Lesson #1 is to compose the scene. A forensic photographer needs to set the correct perspective, angle, and extent of the photograph in order to **communicate** the scene **accurately** to the court in a **minimum** number of photographs. The ability to do this is what sets apart a professional from an amateur photographer.

In the same way that a forensic photographer needs to understand the nature of the crime and the criminal in order to compose the scene properly, market analytics needs to understand the CONTEXT of the industry and specific questions being asked by the business. This information serves as the basis for the hypothesis that is being tested in the analysis.

For those of you who are familiar with Tom Khabaza's 9 Laws of Data Mining: we could say that "Composing the Scene" is just another way to state the first and second law of data mining- **Business objectives are the origin of every solution** and **Business knowledge is central to every step of the process**

Composing the Convenience Store Scene

- According to NACS, there are about 152,794 c-stores in the US.
- 71% of total sales is from fuel, but that is only 36% of profit dollars.
- 50% of all single serve water bottles and 45% of sports drinks are sold by c-stores.
- C-stores check 4.5 million IDs every day for tobacco and alcohol sales. (TSA only does 1.8 million ID checks per day.)

Let's compose the scene for convenience stores. Trade group associations are a good source for background information about an industry. For convenience stores, the National Association of Convenience Stores is one of the best sources of information.

152,794 c-stores compared to only 55,872 building materials, hardware or garden supply stores and about 58,000 department stores. There are more than 2.5 x as many c-stores as these larger store formats but there are 284,992 apparel and accessory stores and over 1 million eating and drinking places in the US. Clearly, c-stores do not have the potentially "unique" inventory that prompts the large number of apparel or accessory stores, so they fall in the middle of the spectrum when it comes to merchandising.

Composing the Convenience Store Scene

- Gross margins are comparable to grocery stores – around 27%; net profit margin is only about 1%.
- C-stores do about 1,130 transactions per day and average 7,397 transactions per week.
- Heavy shoppers only spend a little over $8 per transactions in the store.
- If an average store does $3 million annually, the net profit would be about $6,000 per week.

1,130 transactions per day averages to 47 people/cars per hour for 24 hours operations or 94 people/cars per hour for 12 hour operations. There are only 1440 minutes in a day—a convenience store averages almost 1 sale every single minute!

Speed and volume are the keys to their success . . . And from a customer's perspective, it's all about convenience and THAT is tied to LOCATION more than to any other type of retail business.

Composing the Convenience Store Scene

- Convenience store prices are higher due to:
  - Higher wholesale costs for smaller deliveries
  - Higher utility expenses per square foot
  - Higher real estate costs for corner lots and locations along commuter routes
- Industry faces competitive challenges from
  - QSR, sub shops, coffee shops
  - Drug and grocery stores
  - Warehouse and club stores
  - Mass merchandising stores

In order to analyze trade areas properly, we need to understand where the pain points are in the industry. The C-Store space is different than other retailers because it faces higher costs and must charge higher prices than its competitors. It has also faced a significant erosion of its customer base due to other retailers offering similar value propositions in specific parts of the business.

Composing the Convenience Store Scene

- **The Challenge:** Understand the current trade areas and identify the best growth strategies for the network.
- NACS has defined 4 Growth Strategies
  - Fresh Value Fast (QSR concept)
  - Family Time (Yogurt concept)
  - My Time (Dollar store concept)
  - Female Friendly (Coffee shop, snack break concept)

The challenge for C-stores is to understand their current trade areas and identify the best growth strategies. NACS has defined 4 different growth strategies to help their members: Fresh Value Fast, Family Time, My Time and Female Friendly. But it is probably not possible to be successful with all 4 strategies in a single location.

Disclaimer

- The data used for this presentation is from Esri Business Analyst software with 2014 data.
- Circle K is not a client of 4CTechnologies or the 4CGeoWorks division and no proprietary or confidential data is being disclosed.
- Store locations were not verified due to time considerations.
- The data presented is for demonstration/instructional purposes only.

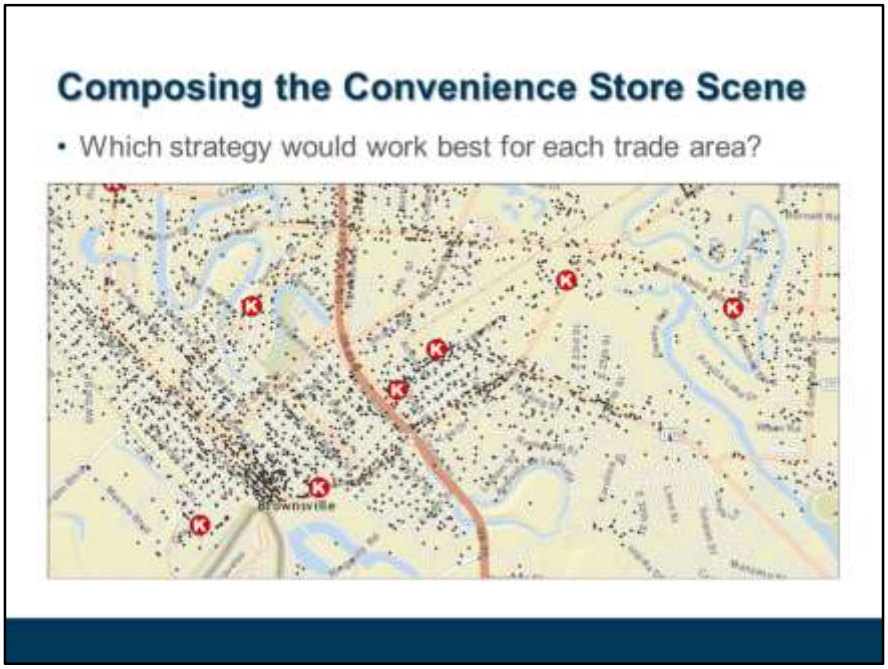
I sign NDAs or confidentiality agreements with almost all my clients, so I'm not able to discuss my project work. That said, I believe very strongly that GIS professionals need to share their expertise and experiences so that we can improve our analytics.

Composing the Convenience Store Scene

- Circle K has over 3,300 stores in the US and more than 4,000 locations in other countries.
- March 16, 2015, the Circle K's parent company, Alimentation Couche-Tarde, purchased about 1,500 c-stores from The Pantry, which is mostly branded as Kangaroo Express.

Berk, Brian. "Couche-Tard Hops on Kangaroo Express Integration" Convenience Store News, 17 March 2015, 05:28 pm. http://www.csnews.com/industry-news-and-trends/corporate-store-operations/couche-tard-already-integrating-kangaroo-express-locations?cc=3 Web. 20 April 2015.

I selected Circle K for this analysis example because it has a large install base of company owned and franchisee owned locations.



The preliminary data preparation steps “composed the scene” for my analysis. The goal is to identify which strategy to employ for each location and to do this in a few days and not weeks. I needed to generate the insights quickly on a national basis. Simply putting the “dots on the map” does not solve our question of which strategy would work best for each store. Now what? How could I create trade areas to answer my question? This brings us to the second lesson from forensic science.

One of the most interesting things that Nick taught me is that, *for a lot of his work*, the camera is irrelevant. The important thing is the LIGHT—where it is, where it is going, and what we can see in an image through the use of different wavelengths. Yes, the object being photographed is important and the camera is important as a tool, but the LIGHT is what reveals the evidence. Nick talks about the fact that forensic photography is a **mental abstraction process** and it is the light that allows the photographer to identify and highlight **what is relevant and what is not**.

We need to create a trade area that is relevant and valid for the business question. The "light" is our tool that we can bend and shine on the trade area in order to define the relevant market for a store. We have many tools to define a trade area, but which one is the right tool? Is it rings, drive time areas, custom drawn areas? What if we have hundreds or thousands of sites? Is it realistic to try to create thousands of hand drawn trade areas? If we have customer data, we can define the trade area by a percent of sales or a percent of customers. But what if we do not have customer data, what do we do then? How can we get the most accurate picture of a market with the information that we have?

From my "composing the scene" analysis, I knew that I needed to account for both residential, employment, and traveler based customers. **And** I knew that I needed to model the customer base from the creation of trade areas.

From Tom Khabaza's 9 Laws of Data Mining, we are essentially covering Law #3 and #4:
Law #3: "Data Preparation Law" - Data preparation is more than half of every data mining process
Law #4: "No Free Lunch for the Data Miner" - The right model for a given application can only be discovered by experiment

Simply drawing a ring around each location was not going to give an accurate picture of the diversity of the trade area sizes.

Drive time areas take a long time to create and removing the overlap sometimes leads to geometry errors.

I tried non-overlapping rings and equal competition areas (Thiessen polygons). Unfortunately, that did not work. The non-overlapping rings didn't provide a good way to calculate the market size for convenience stores along interstates, in rural areas, or in downtown (walkable) areas where many of the customers didn't live there and were only buying from the stores because they worked nearby. The equal competition polygons had much the same problem and they didn't account for the shape of the trade area where people lived or worked.

The Wrong "Light" for C-Store Trade Areas

- Customer data is not available to create customer derived trade areas.
- Manual trade areas are the most accurate, but not worth the time investment for over 2,700 locations.

I didn't have customer data to create customer derived trade areas. I didn't have time to manually create accurate trade areas based on multiple and diverse site characteristics.

At this point, I stopped and "got the light right".

I stopped and looked at the data that I had and asked myself how I could transform it into the data that I needed. I realized that I needed to use the customer derived trade area tool, but I didn't have customer data so I decided to create some pseudo customer data from what I already had at my finger tips in the Business Analyst.

Creating customer derived trade areas is a great way to understand a market. The pseudo customer layer would be created to account for both residential and daytime population. What data did I have to do this? I had almost 24 million business points from D&B and I had about 6 million census block points. The D&B data had a field that identified the number of employees (my daytime population number) and the census block points had total population. I could combine these to develop a data variable for the "demand" side of the equation.

I knew that the customer derived trade area tool needs a single point layer as an input, so I would have to merged the business points and census blocks into a single point layer.

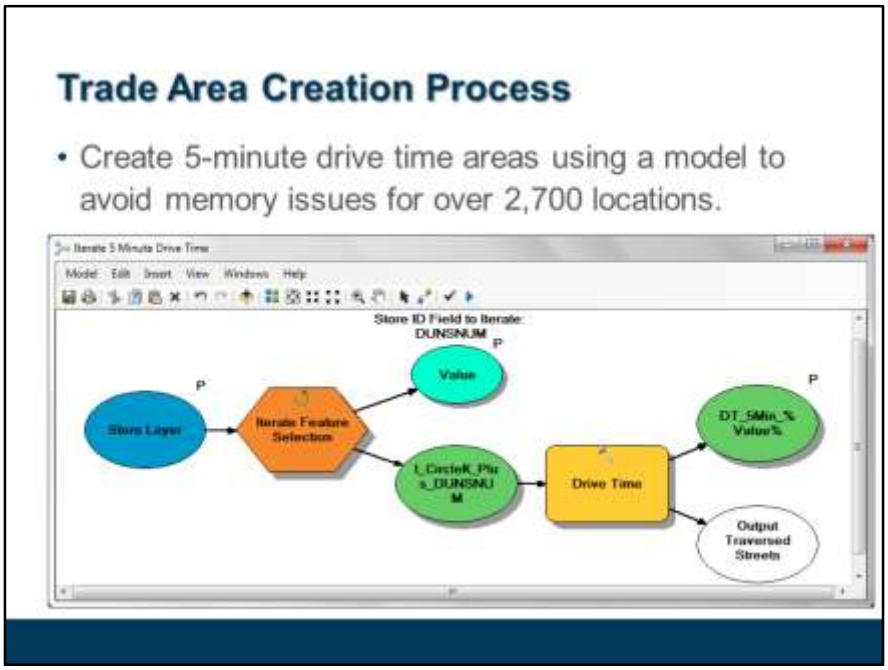
I knew that trying to process 30 million points was not going to be feasible or even a good way to model the trade areas.

I could select only those block points and business points that were within 2 miles or a 5 minute drive time of my store locations.

The Right "Light" for C-Store Trade Areas

- **Solution:** Use the data that I have to create a "pseudo" customer layer.
- Select points within a 5-minute drive time area.
- Assign the points to the closest store to remove any overlap.
- Use the "Customer Derived Trade Area" tool to create areas that are a good representation (abstraction) of a trade area

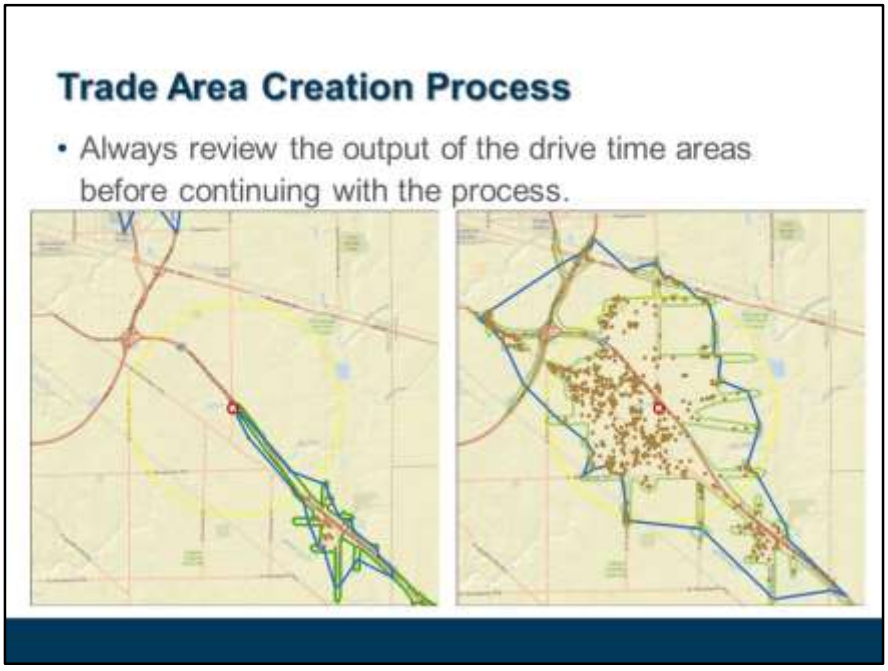
So my solution was to use my "pseudo" customer points and my store points, select points within a 5 minute drive time or 2-mile ring and then assign the points to the closest store to remove any overlap in the trade areas. Once I had done that, it was a cinch to create non-overlapping trade areas for each store that varied in size based on the proximity of neighboring stores as well as my pre-defined distance cut-off value.



Using a model is the best way to calculate a large number of drive time areas to avoid memory issues.

Trade Area Creation Process

- Creating customer derived trade areas for 2,703 locations and 3.8 million customer points only takes about **6 minutes** if you have an attribute index on your Store ID field. Without an index, the process takes around 30 hours.
- If you have 10+ million points, it may help to run the processes in ArcCatalog instead of ArcMap.
- If your computer lacks memory or processing power, you can also try using an iterated model approach and then merge the outputs.



The drive time area is shown in green and the customer derived trade area is shown in green. The area on the left side is obviously not a good representation of the trade area. The area of the right is the “fixed” drive time.

Trade Area Creation Process

- Location matters for drive time analysis. Even a small difference in location can have an effect on the resulting drive time.

The “fix” was simply moving the point a couple dozen yards so that it was closer to the two-way road instead of the to the one-way highway ramp.

Trade Area Creation Process

- For very rural areas like Princeton, Maine, the 5-minute drive time area is not appropriate.

Trade Area Creation Process

- For Chatsworth, GA, the combination of the 5-Minute drive time area and the assignment of the demand points to the closest store yielded a strange trade area. In this case, using a selection of points within either the drive time or the ring area would have given better results.

Lesson #3 is to build a **sequential pattern of search**. Nick Marsh mentions in the video that forensic photographers use a process of looking at the evidence using multiple types of light. Natural light produces some patterns. Adding bright spotlights may yield a better view. Ultraviolet light is used in some cases. Green lasers at other times. Often the light can uncover patterns in the evidence that are not visible any other way and certainly not without the addition of the special light wavelengths.

Location Analytics is an **iterative process** that is focused on solving problems and achieving business goals. It is a process that, like forensics, is **a sequential pattern of searching** for answers to our business questions.

Our answers are typically found in identifying patterns and as the 5th Law of Data Mining states **- There are always patterns**

You have to have a methodology, process or techniques to evaluate the trade area and determine what it means. So the result was these beautifully shaped trade areas, but my next questions to ask are: Are these shapes a **good** definition of the market? And how do I match the stores to a growth strategy?

Not all c-stores have the same type of customer base—some stores have a neighborhood base. They were not located along a major highway, near hotels, or other types of business that would bring customers from outside the trade area into the store. These neighborhood stores had a good residential base and some also had a good base of daytime population for local businesses. For these stores, the trade areas were a very good approximation of the size and shape of the market.

Other stores are commuter based. These commuter based stores were located near a highway entrance ramp or they were located along a major arterial street. The trade areas for these stores would be influenced somewhat by the local trade area, but the character of the customer base would be more like the customer base for the larger market area rather than the demographics of the immediate neighborhood.

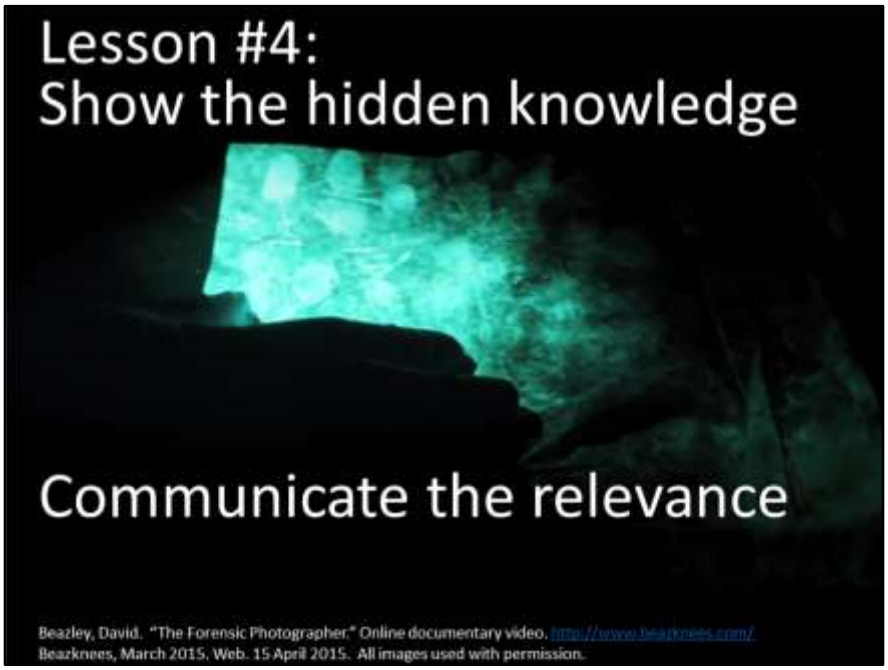
There are other stores that are not only commuter based, but they also benefited from an external base. These were downtown "metro" locations without fuel where the base is a walking instead of

driving customer . . . or locations near amusement parks, hotels, convention centers, and other types of businesses that had a high degree of attraction. These “external base” stores would be in a class all of their own when it came to their merchandising and growth strategy plans.

I identified business demand generators and then used a sequential process of spatial queries and attribute queries to classify the stores. I used a spatial overlay to append demographic data and Tapestry data to look for patterns of density, income, age, and consumer buying patterns. By combining many techniques, I was able to interpret the characteristics of the trade areas and substantiate my assumptions for what would sell well in different types of stores. I could rank the stores and understand which trade areas had the most potential and which had the least. By defining benchmarks, I could group the stores into different classes so that a marketing and growth strategy could be built through a small set of themes that could be implemented for groups of stores rather than trying to develop thousands of specific strategies for individual locations.

Searching for Trade Area Patterns

- Use proximity analysis (Summarize) to identify distances to the closest highway exit, major road, number of hotels, schools, large businesses, etc.
- Append data (spatial overlay) to the trade areas.
- Rank the appended data and proximity results
- Create indexes to compare data
- Use grouping analysis to identify different types of trade areas
- Compare index values among different groups



Lesson #4 is perhaps the most important. It's about a mindset—an attitude—and the goal that guides the entire process of forensic photography as well as GIS analysis. We have to keep an open mind and even though it is trite to say—we have to "think outside the box" in order to uncover the hidden knowledge about the trade areas.

It is not enough for a forensic photographer to uncover the evidence in a photograph, it is often necessary for the forensic photographer to **explain the significance of the finding**. As GIS professionals, we have to be able to distill the knowledge that we have gained from an analysis and present it to others in a way that is easy for others to understand the importance of our findings.

The 6th Law of Data Mining is **"The Insight Law" - Data mining amplifies perception in the business domain** Insight is the act or result of **perceiving the value of the information** for a business. And that is our ultimate goal—insight and recommendations for action to improve the business.

Although it is difficult to interpret this data on a national level through a simple point map on a small slide, the map shows the results of the iterative analysis process of assigning each store's trade area to a growth strategy. You can use the Data Driven Pages toolbar to create a series of maps at the local market level to more clearly see the patterns for each market.

Having this foundation, allows you to think creatively about how to solve the problem of creating a trade area and how to select the relevant data that will formulate an accurate picture of the market that each store will serve. The demographic and consumer buying patterns are there—if we only know how to shine the right light on the data, compose the scene correctly, build a sequential pattern of search and find the hidden knowledge.

14 Technical Tips & Tricks

1. Use ModelBuilder with an iterator to create drive time areas for a large number of stores (and then merge results).
2. Running models in ArcCatalog instead of ArcMap is sometimes much faster.
3. Make sure that there are no duplicate store locations before using this methodology.
4. Create an attribute index for the StoreID in the pseudo customer layer to improve processing time (6 minutes versus 30 hours)

14 Technical Tips & Tricks

5. Send model output to a new file geodatabase when you have thousands of output files.
6. Use "Detailed Drive Time" for a more conservative trade area size. "Standard Drive Time" is faster but creates a larger trade area size.
7. Use the Merge tool in ArcToolbox and not a model or the Geoprocessing menu for faster processing.
8. Review drive time areas before using them to select pseudo customer points.
9. Consider grouping stores into 2 different drive time areas- 5 minutes for most and 10 minutes for rural locations.

14 Technical Tips & Tricks

10. Hide unnecessary fields when merging business points and block points so that data processes faster.
11. Add a UniqueID field to the pseudo customer layer based on the ObjectID field.
12. Use Spatial Join in ArcToolbox for the fastest assignment of the customers to the closest store.
13. Run the Repair Geometry tool on the trade areas to ensure polygons have the correct geometry.
14. Esri recommends the "Detailed with Smoothing" option for customer derived trade areas, but "Detailed" worked better in this case. Amoeba and Simple also produced less desirable areas.