Applying census tract data to a campaign: a low-cost solution

For applying demographics, ZIP code or census tract?

It is standard practice in direct mail campaigns to consider demographics at the _ZIP code_ level. After all, that is the "geographic" unit common to both mailing lists and most sources of demographic data. However, as we will see:

ZIP-code-level data is often significantly inaccurate.

Census tracts are the wise choice for a demographic unit.

To incorporate census tract information in direct mail campaigns:

Using a packaged solution can cost over $3000 _per year_

… but I propose a low-cost solution, which we will see.

This deck will look at why:

- **Census tract is a much more granular, accurate and useful geographic unit**
- **It can be implemented with a solution that is easy to execute and inexpensive**

This will result in more cost-effective direct mail campaigns.

The evidence for such an approach will be documented throughout this deck. Also, all the graphics illustrating how such a solution could be implemented are actual screen shots of me executing the solution I propose.

In planning a real direct mail campaign, many demographic data sets should be considered. For the purposes of this report and in the interest of simplicity, we will consider only one demographic: homes built during the 1950s.

Applying census tract data to a campaign: a low-cost solution

What is a *census tract*? In a word: **neighborhood**. The Census Bureau says census tracts are:

"[d]esigned to be relatively homogeneous units with respect to **population characteristics, economic status, and living conditions** at the time they are established… [They] generally contain between 1,000 and 8,000 people, with an optimum size of 4,000 people. Census tract boundaries are delineated with the intention of being stable over many decades, **so they generally follow relatively permanent visible features**… [my emphasis]." *

The map, above, shows some neighborhoods in downtown Minneapolis. The Warehouse District (peach) is census tract 35.01. Loring Park (bottom, purple) consists of census tract 1052 down to 15th Street --- then census tract 1056 south of 15th.

If you have walked around Loring Park, you would notice that 15th Street is clearly the boundary between two different household income groups. The census tract boundary reflects this.

* "Decennial Management Division Glossary", U.S. Census Bureau, undated (http://www.census.gov/dmd/www/glossary.html)

Applying census tract data to a campaign: a low-cost solution

Why not use ZIP code-based demographic data? After all, they are available from both the U.S. Census Bureau and commercial vendors and they are easily applied to prospective mailing lists.

A University of Missouri report explains the reasons why (all emphasis mine):

"We see maps of ZIP codes in telephone books and from commercial vendors that make us think of them as spatially defined areas with precise boundaries, similar to counties. But, from the perspective of the agency that defines them, the U.S. Postal Service, ***ZIP codes are not and never have been such spatial entities***.

They are simply categories for grouping mailing addresses. ...[T]he actual ***definition of a ZIP code 'boundary' is quite fuzzy at best***, and a purely extrapolated guess... at worst. If you have an application that requires extreme geographic precision, especially in sparsely populated areas, then you need to avoid using ZIP codes."[1]

Because of this "fuzziness", the Census Bureau's geographic units (census tracts, blocks, etc.) cannot match up with ZIP code "boundaries." The U.S. Postal Service admits that a ZIP code, even a nine-digit ZIP+4 code, can often be "matched with more than one census block" because the ZIP "was in more than one census block" or there was "[a]mbiguous data" on which to base a census tract/ZIP translation.[2]

Census data is collected and reported according to Census Bureau-defined geographic units. It makes sense to apply that data ***by those units***. Otherwise, our view will be distorted when we force census information into ZIP codes. We will see these bad effects in our next slide.

1 "Tools and Resources Related to U.S. ZIP Codes, Version 2.1", John Blodgett (Office of Social and Economic Data Analysis, University of Missouri), June 2006 (http://mcdc2.missouri.edu/webrepts/geography/ZIP.resources.html).
2 "Topological Integrated Geographic Encoded Referencing", U.S. Postal Service, 2007 (http://ribbs.usps.gov/files/addressing/pubs/tigerzip.pdf).

Applying census tract data to a campaign: a low-cost solution

Step 1: Determine majority ZIP Code for census block

Census Tract 12 – Block 1102

Street Address		ZIP Code
1	Dee St	22345
3	Dee St	22345
9	Dee St	22345
10	Jullian Dr	22345
105	May Rd	22346

Majority ZIP: 22345

Dee St · May Rd · Outer Creek · 1 · 3 · 9 · 105 · Block 1102 ZCTA 22345 · 10 · Jullian Dr
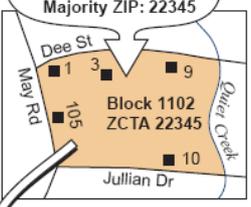
Step 2: Extend ZCTAs into blocks not assigned ZCTA

ZCTA 22345 · ZCTA 22346 · Outer River · ZCTA 22344
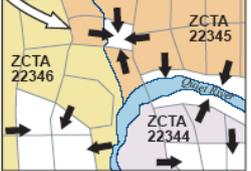
Step 3: Run edits and review results of process

ZCTA 22346 · ZCTA 22345 · ZCTA 223HH · ZCTA 22344 · Outer

Figure 6: **Steps in creating ZCTAs**

Remember that there are ___no official ZIP code boundaries___: they are merely a best guess by whomever is drawing the map.

The Census Bureau's method of matching is shown on the left. The method is to create a new geographic unit called a "ZIP code tabulation area" (ZCTA). ZCTAs are defined by the majority ZIP code of a census block, which is the demographic unit one level smaller than a census tract.[1]

The problem with this approach is illustrated in the U.S. Census Bureau graphic below. [2]

ZCTAs may reduce the boundary matching problem, but they do not eliminate it.

    ZIP code "boundary"

    Census block boundary

Figure 2: **Differences between ZIP Codes and ZCTAs for a neighborhood**

Meyer Creek · ZCTA 23544 · Block 1112 · Kansas RR · 511 · 11th St · Western · 501 · 505 · 601 605 · ZIP Code 23584 · 504 · Fir St · 600 · 606 · Block 1114 · Block 1113 · ZIP Code 23544 · 23546 · 711 713 · 803 · 807 · May School · 714 716 · Pine Rd · 810 · Block 1115 · MNN Inc. · ZCTA 23546 · ZIP Code 23561

ZCTAs represent the majority ZIP Code for MAF addresses within a tabulation block.

This ZIP Code is for the P.O. Box that serves 611 Fir St.

ZIP Codes are based on street deliveries, not block boundaries.

This ZIP Code is for the firm.

1 "Census 2000 ZCTAs ZIP Code Tabulation Areas Technical Documentation", U.S. Bureau of the Census, undated (http://www.census.gov/geo/ZCTA/zcta_tech_doc.pdf).
2 Same as above.

Applying census tract data to a campaign: a low-cost solution

Distortion in ZIP code demographics also occurs from neighborhoods with such high or low numbers that they skew our perception of the ZIP code. Consider the two maps below. Both show 1950s housing unit counts for the same area of the Twin Cities. The version on the left displays counts by *census tract*, the map on the right by ZIP code. Because census tracts represent 4,000 households on average while ZIP codes represent 10,000 (150% more), I adjusted the scales accordingly.

Number of 1950s homes by CENSUS TRACT

Number of 1950s homes by ZIP CODE
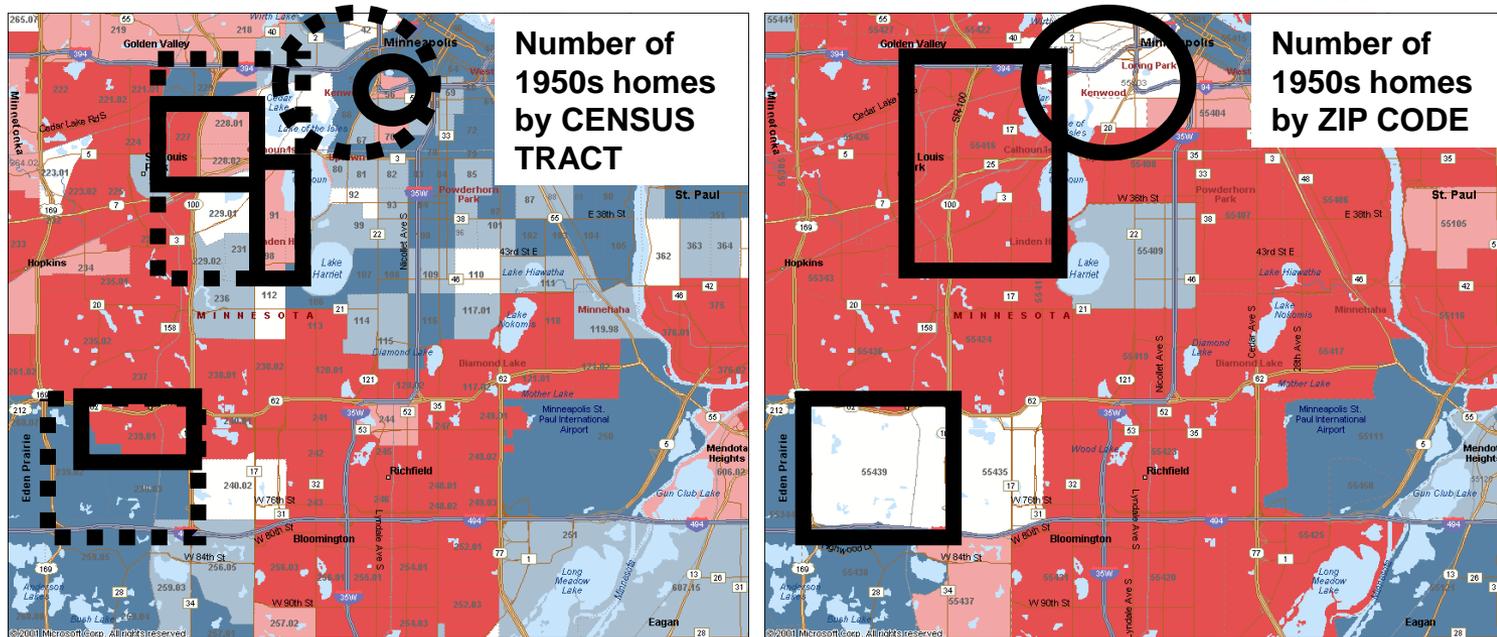
By census tract

🟥	>400
🟧	301 to 400
⬜	201 to 300
🟦	100 to 200
🟦	< 100

A few high-count neighborhoods (solid lines, left) inflate the importance of the larger ZIP code to which they "belong" (right). Why not simply mail to high-count neighborhoods and save on postage?

By ZIP code

🟥	>1000
🟧	751 to 1,000
⬜	501 to 750
🟦	250 to 500
🟦	<250

Applying census tract data to a campaign: a low-cost solution

Now that we have seen that ZIP code demographics should not be used, how can we incorporate census tracts in a direct mail campaign? Below is the outline of how. In the slides that follow we will look at each step in greater detail.

Implementing census tract data in your campaign planning

STEP 1
Use MS MapPoint to extract the census tract demographics for areas we are considering for our campaign. Decile or otherwise prioritize census tracts accordingly.

STEP 2
Use the U.S. Census Bureau website _and_ web scraping software for an automated search & extraction of census tracts for each address in our potential mailing list.

STEP 3
Once you have census tracts identified for every address in your potential mailing list, you can match them to the results of Step 1. This will let you execute a well-targeted direct marketing campaign.
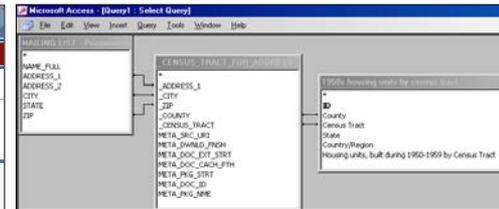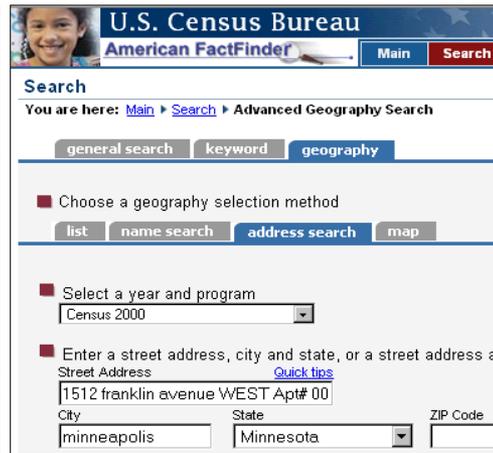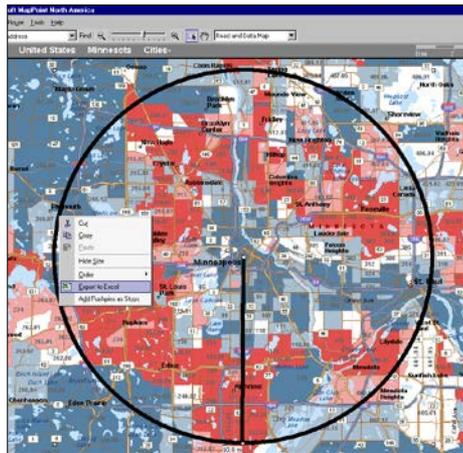
WillBeauchemin.com

Applying census tract data to a campaign: a low-cost solution

Step 1: Getting census tract demographics.

While the Census Bureau's website provides this data for free, its format is awkward for extraction. Microsoft MapPoint is a much simpler solution.

At left is an MS MapPoint map displaying census tracts by number of 1950s housing units. I can extract this data from MapPoint by selecting an area on the map and exporting to Excel (see red arrow, left).

In this example, I have selected a 10-mile radius around Minneapolis. The process would be much the same for an entire major metropolitan area, an entire state, or a regional territory. A circle is not necessary: you can just as easily outline the borders of a state or sales territory.
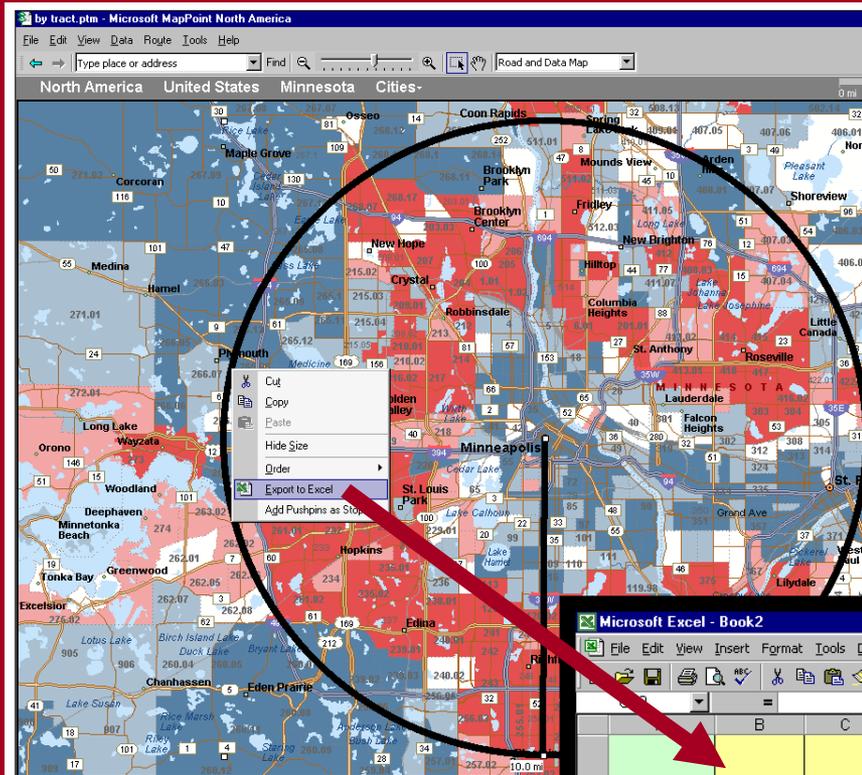
Once extracted, we can decile or otherwise prioritize census tracts by their demographic data.

WillBeauchemin.com

Applying census tract data to a campaign: a low-cost solution

MS MapPoint provides both U.S. Census data and commercially-gathered demographic data.

An added benefit of using MS MapPoint to extract census tract information is the large amount of commercially-gathered demographic data (non-census) that comes bundled with the software. Here is just a small sample of the demographics that MS MapPoint 2006 provides (and the as-of year of the data):

- Median disposable income (2004)
- Population, Ages 45-54 (2004)
- Households, Income $100,000-$149,999 (2004)
- Total crime (2004)
- Average household expense, Entertainment (2004)
- Commute, 45 minutes or more (2004)
- Households, Next vehicle purchase SUV (2003)
- Adults who purchased books in last year (2003)
- Adults with home improvement loan (2003)
- Adults who purchased from catalog in last year (2003)
- Households with satellite dish (2003)
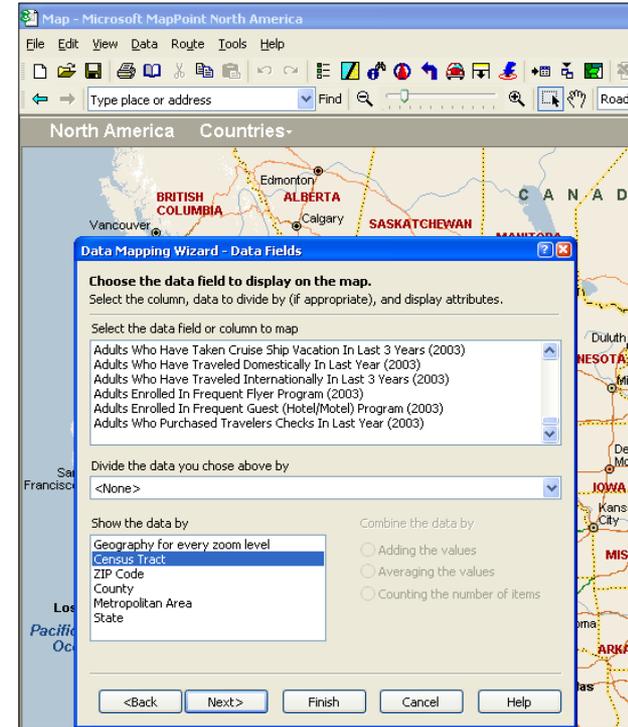
Preview MS MapPoint online or download a free trial at
http://www.microsoft.com/mappoint/products/2006/trial.mspx

WillBeauchemin.com

Applying census tract data to a campaign: a low-cost solution

Step 2: Identifying the census tract of each address.

To tie our census data to our potential mailing list, we need to identify the census tract of each of our addresses. The U.S. Census Bureau's website has a tool to do this.

The tool is flexible in handling variations of addresses. In the example, right, note how the spelling of "avenue" did not prevent the tool from finding a match --- nor did the placement and spelling of the directional "WEST" and the presence of an apartment number (which is ignored). No special address standardization would be needed prior to running a mailing list through this lookup tool.

A match is found.

Performing lookups with this tool by hand would be impractical for large mailing lists. However, by using special "web scraping" software, we can easily program both the search and the extraction of this information. We will see how this is done in the next slide.

WillBeauchemin.com

Applying census tract data to a campaign: a low-cost solution

The software lets you extract web page data to specific fields in a database table. You define which data on the web page that should be based on the particular text (HTML, XML, preceding/following text) that surrounds each datum you want. This flags the datum for extraction.

Step 2 (*cont.*)

We automate the task of looking up census tracts by using "web scraping" software, such as Web Scraper Plus+. Shown left is the "datapage" I created in Web Scraper Plus+ to do this.

After feeding my mailing list into the software's automation tool, the datapage performed the lookups on the Census Bureau website and extracted the results to an MS Access table.

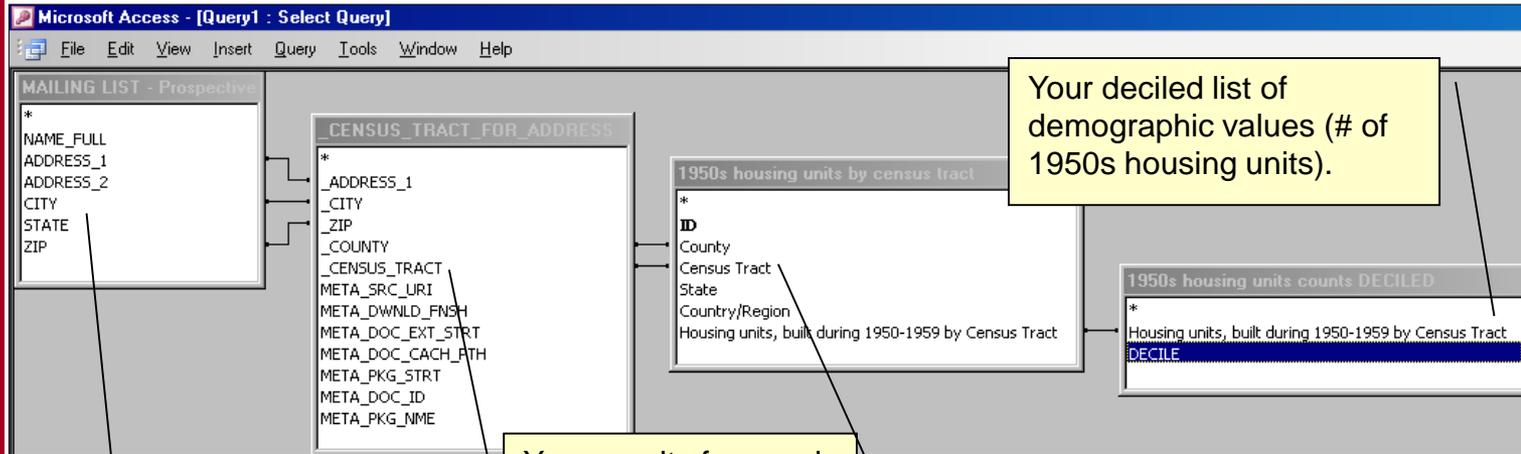Web Scraper Plus+ works with almost any database program. It can also export to MS Excel.

Applying census tract data to a campaign: a low-cost solution

Step 3: Link your results to your deciled/prioritized table of census tracts.

Take your database table of extracted census tract-address data and link it to your original prospective mailing list. Now you have the full mailing address (householder's name included) together with the census tract.

Now link those results to your prioritized list of County/Census Tract # combinations. You can now create a targeted mailing list.

Your deciled list of demographic values (# of 1950s housing units).

Your prospective mailing list

Your results from web scraping the Census Bureau website for census tract matching to addresses.

Your census tract demographics you exported from MS MapPoint (saved to MS Excel and imported to MS Access).

Applying census tract data to a campaign: a low-cost solution

Assuming that any marketing department would already have relational database software, only two other software products would need to be purchased for the solution I propose:

Microsoft MapPoint 2006
Cost: $299 (single license)
Learn more: http://www.microsoft.com/mappoint

Velocityscape Web Scraper Plus+
Cost: $749.00 (single license)
Learn more: http://www.velocityscape.com/Products/WebScraperPlus.aspx

Another key component of the solution is FREE:

U.S. Census Bureau
Cost: Free
Census tract lookup by address:
http://factfinder.census.gov/servlet/AGSGeoAddressServlet?_lang=en&_programYear=50&_treeId=420

All prices as of 2/24/2008

Applying census tract data to a campaign: a low-cost solution

What I have proposed is _much less_ expensive than a packaged solution.

Consider the software product Mailers +4 Pro from the company Melissa Data. With its add-on product GeoCoder, it can take a mailing list and append the corresponding census tract identifier. It can also link to Census Bureau demographic data.

To use this commercial solution, you must buy their base product **_plus_** one-year-only licenses to two sets of data:

> MAILERS+4 Pro Single….……….$495 (one-time cost)
> MAILERS+4 Pro 1-Year Sub….$1,695 (annual cost)
> GeoCoder 1yr Sub….………..….$1,490 (annual cost) *

Compare the $3,185 **_annual_** cost of the packaged solution to the $1048 **_one-time_** cost of the solution I propose.

* All prices as of 2/24/2008
http://www.melissadata.com/geocoder/geocoder.html

Applying census tract data to a campaign: a low-cost solution

Use census tracts to plan your campaign for greater cost effectiveness

By getting more granular through census tracts, you can better target households with your more expensive mail pieces and better offers.

There is a demographic unit smaller than a census tract (it is called a "block"). However, the number of such small units becomes unwieldy when you plan anything larger than a metro-area campaign.

Census tract is the practical choice --- and the better choice compared to unreliable ZIP code demographics.

While commercial solutions are available, there is a less expensive option

With geographic marketing software such as MS MapPoint (and the census and commercial information that is bundled with it), census tract information can be easily extracted in tabular form. You can then prioritize the census tracts by their demographics. Combine different demographics for a multi-dimensional approach to planning your campaign.

With web scraping software, such as Velocityscape's Web Scraper Plus+, you can automate the appending of census tract I.D.s to your mailing list.

I've already created the automation tool. Contact me. I'll send it to you.

If you own a license for Velocityscape's Web Scraper Plus+, contact me. I will email you a copy of the datapage that automates the census tract-address matching.

Questions? Contact me at Himself@WillBeauchemin.com