DO LOTTERIES EXPLOIT THE POOR?

UNZIP FOR A BETTER VIEW.

STEPHEN WADE, RESEARCH AND DEVELOPMENT MANAGER, WASHINGTON'S LOTTERY AND PRINCIPAL, LOTTERY MANAGEMENT CONSULTING, LLC.

The assertion "lotteries exploit the poor" is so frequently repeated that people may come to regard it as fact, or at least as received wisdom. When I ask people how they know this, they can generally refer to some more-or-less current piece of journalism they have read. When I look at these sources, I find that they sometimes present as investigative journalism, where the writer has obtained data from a lottery and used it to provide supporting evidence – "look, the lottery's own records prove that it exploits the poor."

My purpose in this article is show how the lottery's records are typically used to support this type of claim. I take some satisfaction in exposing how convenient and unquestioned assumptions can lead to conclusions that fall apart on closer examination. The level of public discourse may now be too low to hope to revise public opinion with reasoning, but I will aim to conclude with something suitable for a 5-second "sound bite" – perhaps enough for current appetites.

I will show you how superficial facts and easy assumptions support a conclusion that Washington's Lottery exploits the poor. Then I will show how, with better information and better analysis, this support disappears.

It all starts with a public information request, typically for lottery ticket sales information tied to geographic locations (like the address of the retailer). Washington's lottery has this information readily available, and turns it over promptly.

What would happen if the analyst asked for lottery sales tied to individual players? Someone from the lottery might take time to explain that such records do not generally exist. In most cases, lottery tickets are purchased with cash by players of unknown identity. The most we know is where and when tickets are purchased, and that is what we share.

The analyst, though, wants to relate the location of the sales to the income of the buyer. The key to this is census data. The U.S. Census Bureau conducts a massive program of household interviews every decade, with American Community surveys in between. The fundamental data are reportable by something called a census tract, which is different from a zip code or city or county. However, data can be aggregated by zip code, the most common geo-locator in business records. Extracts of census data by 5-digit zip code are widely available, and commonly represent measures of population, average income, housing, employment, and so on.

So, the analyst expects to know all about the demographics of people who buy lottery, from knowing the zip code of the location where sales are made. And usually the first thing the analyst will do is calculate a rate of consumption for lottery products: annual dollars per capita, for instance, or weekly dollars per household.

Note that there is a special poignancy in the idea of "household spending on lottery per week" – the units help us imagine people sitting around a kitchen table, reaching into a money jar to see what might go for food and what might go for lottery tickets that week, a scene to be repeated every Friday night. In fact, this scenario is not usually supported by the only kind of data that would test it: household-level interview data. This kind of data requires a highly structured and expensive program of surveys. In

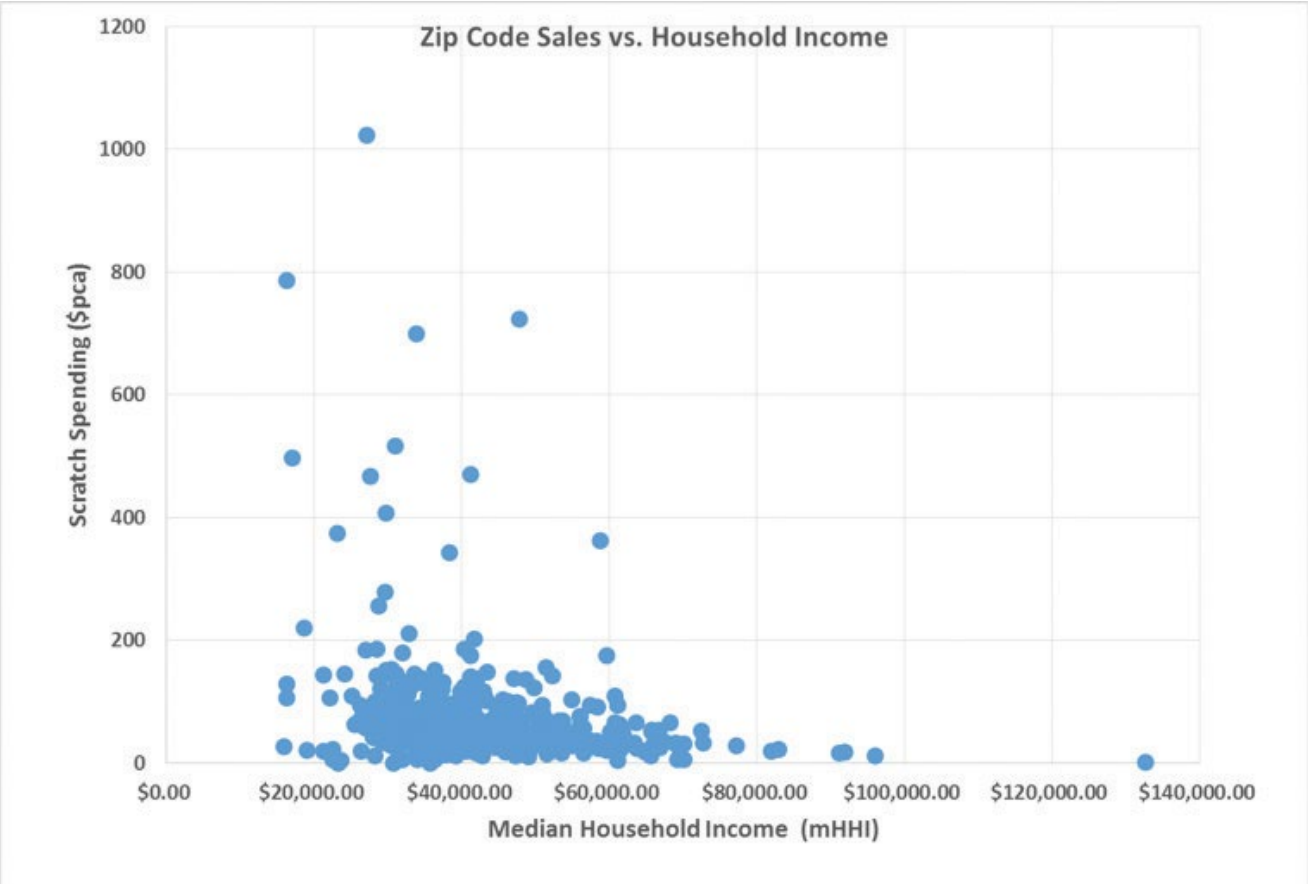
the usual story about the lottery, no one living in a zip code was questioned about lottery play to come up with "household spending on lottery." People and households were counted within the zip code, and the lottery revealed what was sold by its retailers that are located within the zip code. Even though these human-scale rates may be spoken of as averages for the zip code, they are merely calculations.

From a purely quantitative point of view, this might be reasonable if the zip code were an island unto itself. In fact, one of the charms of Washington State is that some of our zip codes are islands unto themselves – the mail truck arrives by boat. However even on these islands, people do come and go every day. And when the zip code is not an island, but a zone linked by a continuous web of roads connecting a population numbering in the millions, the notion that economic transfers of any kind occur primarily within the zip code is clearly not justified.

But it surely is convenient to assume that people buy lottery where they live. And consequently, this assumption is frequently used, not only outside but even inside the lottery.

So to recap, the easiest thing is to assume that the sales made at a retail location are made to people who live in that same zip code. Having made this assumption, the analyst can quickly access two kinds of denominators that are used (with lottery sales as the numerator) to speak of "rates of sales:" the count of people and the number of households. Going further, the census data offer many other items that speak to living conditions within the zip code: income per capita or per household, rate of unemployment, household size, and so on. It is a simple matter to look across zip codes and see whether zip codes with low income measures have high lottery consumption. For instance, in this scatter plot of per capita Scratch sales vs. median household income (mHHI), there are some areas with extremely high per capita sales and low mHHI, in the upper left quadrant. These outliers attract the eye, and provide visual support for a "lottery exploits the poor" story. Sometimes, journalists just call out these extreme cases to support their story line.

There are multiple problems with this analysis. A very fundamental problem, seen in many areas besides lottery, is in the math: measures that depend on dividing one number by another (like lottery sales per household) can take on extremely high values when the denominator is small. Zip codes are defined for their utility in handling the mail. They take account of geography; they do not all serve a similar number of residents. A zip code with a few hundred people is far more likely to produce extreme values of many measures than a zip code with tens of thousands of residents.



Income and Lottery Spending by Zip Code Size Class

Population upper limit	Count of zip codes	mHHI	Activations $pca
1000	52	$33,004.25	$193.76
2000	37	$34,550.08	$73.46
4000	57	$40,801.82	$59.53
8000	55	$42,088.13	$61.49
16000	59	$42,701.27	$69.00
32000	116	$49,074.37	$59.63
64000	45	$48,827.11	$58.08

Further, by using the zip code as the unit of analysis, in effect the analyst is "promoting" the underpopulated zip codes to have the same weight as others, even though the others may represent 50 times as many people. This can lead to faulty inferences, as developed below.

First, consider that on an overall, state-wide basis, Washington's Lottery sells only a little more than $60 per capita-year ($pca) in instant tickets. This is not because Washingtonians don't gamble, but because they have great access to Native American gaming that nets about nine times as much as the lottery.

The table above shows, for seven size categories of Washington zip codes, the calculated rate of instant (Scratch) sales per capita. The smallest zip codes have both a rate of Scratch sales about three times higher than the largest zip codes, and per capita income that is only about 2/3 that of the most populated zip codes. These points, taken together, certainly support an impression that the poorest people are buying the most Scratch tickets. The smallest zip codes also have fewer than 1,000 people each, and both extremely low (0) and extremely high (>1,000) rates of sales on a $pca basis. These points do not support a story, and tend not to be remembered.

The impression that the poorest people are buying the most lottery tickets is created largely by giving the 52 zip codes with fewer than 1,000 people (total population about 26,000) about the same "weight" as the 45 zip codes with more than 32,000 people (total population about 2.7 million). Does anyone really think that there are more poor people among the 26 thousand than among the 27 hundred thousand? Using zip codes as in the analysis above magnifies the contribution of the least-representative unit up to 100-fold. In the same vein, treating the 421 zip codes as equally-weighted units of analysis would exaggerate Washington's rate of Scratch consumption by about 1/3 - from about 60 to about 80 $pca. Please note that the distortion introduced by weighting zip codes in this way would be equally objectionable, if it somehow cast the lottery in a positive light.

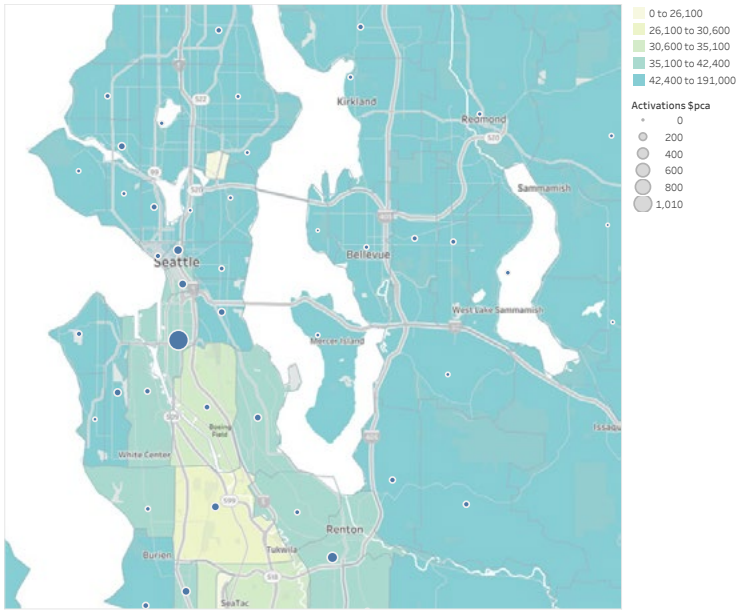
If we really want to know about income and lottery play at an individual level, we must ask individuals. This is something Washington has done consistently in a usage and attitudes survey that has been consistently fielded for many years. That might be a topic for another time.

Even if better data refute the idea that the lottery exploits the poor, seeing an extremely high rate of Scratch consumption in an underpopulated zip code might still be cause for concern. But this brings us to another big problem: the assumption that a sale within a zip code is to a player who lives in that zip code is frequently wrong. In Washington, in fact, it is wrong more often than right. How do I know this? By reference to the only players we really know a lot about- the ones who win prizes big enough to be reported to the Internal Revenue Service ($600 and above). In Washington we usually know both where they live and where they bought the winning ticket. And in my current study of Scratch ticket winners, these two location are in the same 5-digit zip code only 40 percent of the time. This is direct proof that where lottery sales are made does not tell us much about to whom these sales are made. The assumption that the demographics attached to the sales location are the demographics of the players can be wrong- even if this assumption is made as often by lottery analysts as by unfriendly journalists.

In fact, inferring the demographics of the players and where they live is hard. It requires knowing about the players, where they live and what games they have won. Compared to sales data, these records are very sparse. In all our games, many tens of thousands of dollars of sales are recorded for each single win of reportable size. In fact this proportionality between wins and sales is the key to making inferences about the demographics of players- where there is a win, there must have been a sale. Although clearly it is possible for the one and only lottery ticket ever purchased by a player to be a winner, over the long term and over areas of sufficient population, we expect lots of winners where many tickets are purchased, and few winners where few tickets are purchased.

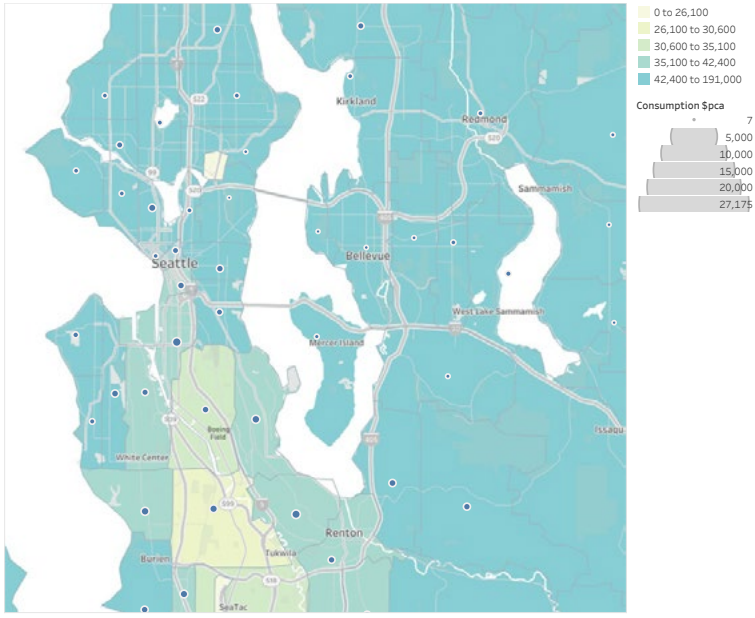
The proportionality between reportable wins and volume of sales differs markedly from one kind of game to another, and it is best to take this into account very precisely. If the volume of sales required to produce a reportable win is very high (as with $1 Scratch games), we cannot expect to map spending on the basis of wins with useful precision. It appears effective to lump instant games together according to price point, and establish the proportionality between all the claimed prizes in some period of time (like five years) and all the sales in that same period of time. Then, everywhere there has been a win, we assume that the proportional amount of spending has happened. We must keep in mind that these data are very "grainy," because it is always possible that because one of the tickets purchased in a zip code was a winner, we infer $10,000 worth of lottery spending where the lucky player really only spent $2. The computations required are only moderately complex.

Activations $pca



Map based on Longitude (generated) and Latitude (generated). Size shows sum of Activations $pca. Details are shown for Zip.

Consumption $pca



Map based on Longitude (generated) and Latitude (generated). Size shows sum of Consumption $pca. Details are shown for Zip.

This is the basis of mapping winners, and inferring purchasers. It is filling in the map with a very broad brush indeed. But when I put purchases into zip codes and compared them to sales, I saw there is a lot of importing and exporting going on. Particularly, sales in some areas amount to much more than local consumption.

In particular, zip code 98134, with the highest rate of Scratch activations (over 1000 $pca), has a moderate rate of Scratch consumption- about 116 $pca. At the last census it was home to 636 people, and some thriving stores, in metropolitan Seattle. The maps shown here contrast two different views of the situation: the first is based on where lottery ticket activations (commonly called "sales") happen, and the second is based on where winners live. Note that "consumption," in contrast to "activations," is remarkably consistent across zip codes.

In summary, I would suggest that any lottery be prepared to share the answer to the following question: "When the address of the winner and the address of the selling retailer are both known, how much of the time do they share the same five-digit zip code?" This is good to know in any case, and some of the time it may serve to prevent a story based on zip codes becoming the sensation of the week. ■