Supercenters, Hamburgers, and Coffee: Using density-equalizing cartograms to display the distribution of Walmarts, McDonalds, and Starbucks in the US

By Steph Abegg

Outline:
- Introduction
- Background
- Graphic Output
- Franchise Trends and Relationships
- Advantages of Cartograms
- Disadvantages of Cartograms
- Conclusions
- Appendix:
  - Data Sources
  - R Code
Introduction

Which state has the most McDonalds? Is Starbucks a west-coast phenomenon? Is Walmart taking over?

Inspired by these questions, the following paper not only presents a unique and playful way of mapping American commercialism, but also uses the analysis as an opportunity to conduct a comprehensive investigation of density-equalizing maps, or "cartograms." First, I discuss the development and methodology of cartograms, including a relatively recent method proposed by Gastner and Newman in 2004. Then, I use Gastner and Newman's method to create elegant cartograms depicting the distribution of Walmarts, McDonalds, and Starbucks in the fifty United States. I examine some of the interesting trends evidenced by these cartograms, and finally I conclude with a discussion of the advantages and disadvantages of cartograms in comparison with other common methods of graphical data presentation.

Background

For statistical data involving estimates for geographic areas, it is natural to want to display the data on a map. Maps are valuable means of graphical display, allowing visualization of trends and patterns that other forms of presentation cannot.

However, there are challenges associated with using cartography to analyze or present statistical data. Apart from needing additional geographic information and the increased complexity of the data presentation, a significant hurdle is that population density is extremely variable. Smaller populations tend to be found in larger geographic regions (such as rural areas), which can lead to misleading visual impressions. Perhaps the best way to visualize data that is affected by spatial characteristics is to actually use spatial characteristics to distinguish the trends on display. This is the idea that inspired cartograms.

Cartograms are maps in which the sizes of geographic regions appear in proportion to their population or some other analogous property. These density-equalizing maps are useful for the representation of census results, election returns, disease incidence, and vital statistics such as the distribution of Walmarts, McDonalds, and Starbucks across the country. Several methods for making cartograms have been proposed over the years, but most of these ideas are inordinately complex or suffer from a lack of readability due to the distortion needed to scale regions and have them still fit together.

In 2004, Gastner and Newman at the University of Michigan proposed a new method of making cartograms. Their method is not only faster and more conceptually simple than previous methods, but also produces useful and easily readable cartograms. Using the relatively straightforward principles of linear diffusion, a cartogram is created from a given population density (or some other analogous property) by allowing the population to "flow away" from high-density areas into low-density ones, until the density is equalized everywhere. Areas shrink and grow and distort to stay connected, producing a cartogram that is in fact unique for a given dataset. (For the formulas and Fourier transforms, Gastner and Newman's 2004 paper is a good reference.) The degree to which the data is binned is important to achieving the desired balance of distortion and recognizability: a very fine level of data binning will cause substantial local distortions, while a coarser level of binning will result in a cartogram with features that are easier to recognize, but gives a less accurate impression of the true population distribution.
The rest of this paper focuses on my use of Gastner and Newman’s method to develop cartograms displaying the distribution of Walmarts, McDonalds, and Starbucks in the fifty United States. The Appendix provides annotated references to the datasets and software I used in this study, as well as provides some examples of code I wrote to analyze and display the density-equalized cartograms in R.

Graphic Output

Cartograms are most often used to show population data, but there is no reason why they need be limited to population. They can be used to show almost any quantity—in my study, the state totals of a Walmarts, McDonalds, or Starbucks. This section uses examples to illustrate the interpretation of cartograms. Provided at the end of this section is the graphical output for the franchise data.

A general interpretation

Below is a cartogram of the US in which the sizes of the states are proportional not to their sheer topographical acreage but to their number of Starbucks coffee shops, which is to some extent a measure of the coffee consumption of each state. States with more Starbucks appear larger than states with fewer Starbucks, regardless of their actual land area. Thus, on such a map, the state of Rhode Island, with 14 Starbucks, appears about 1.5 times larger than the state of Wyoming, with 9 Starbucks, even though Wyoming has 60 times more acreage than Rhode Island.

![Cartogram for the US in which the sizes of the states are proportional to the number of Starbucks. The distortions are indicative of the relative densities of Starbucks in each state. Although not necessary, the color-coding helps distinguish states with similar areas, and the numbers give the state totals.](image)

Figure 1. Cartogram for the US in which the sizes of the states are proportional to the number of Starbucks. The distortions are indicative of the relative densities of Starbucks in each state. Although not necessary, the color-coding helps distinguish states with similar areas, and the numbers give the state totals.
Interpretations can be made of the distortions and color-coding of a cartogram. Using the Starbucks cartogram in Figure 1 as an example, these principles are discussed in the following sections.

**Interpreting distortions**

In Figure 1, the size of each state is proportional to the number of Starbucks coffee shops in the state. In a parallel universe where all states have the same number of Starbucks, the cartogram would be distorted into fifty equivalent areas. If instead all states in this parallel universe have an equal number of Starbucks per square mile, then the cartogram would depict the familiar US map with no distortions. In a sense, the cartogram represents how large each state would need to be to achieve an equal land density of Starbucks throughout the country. Hence, it is the level of distortion, rather than the actual state sizes, that indicates the distribution trends of Starbucks across the country.

A state with a positive distortion—i.e. one that appears "too big," such as New Jersey and California on the Starbucks cartogram—has a higher than average land density of Starbucks. A state with a negative distortion—i.e. one that appears "too small," such as Wyoming or Alaska on the Starbucks cartogram—has a lower than average land density of Starbucks. States that appear "normal" represent areas that have close to the national average density of Starbucks.

**Interpreting colors**

The varying shades of green in the cartogram in Figure 1 represent the varying state totals of Starbucks coffee shops. The color scheme is sequential, with darker greens depicting states with more total Starbucks and lighter greens depicting states with fewer total Starbucks. The ten colors in Figure 1 represent 10% quartile increments in state franchise totals from the minimum (Vermont with 4 Starbucks) to the maximum (California with 2010 Starbucks).

In a sense, the color-coding is redundant since it displays the same information as the areas of the distorted states. States with the same number of Starbucks will be the same size on the cartogram, and also the same color, and visa versa. However, the color-coding helps the eye quickly distinguish states with similar areas, as well as identify regions of the country containing states with similar total numbers of Starbucks.

It is important to note that the coloring does not indicate trends in franchise density, but only represents the total numbers in each state, regardless of state land area. For example, Washington and Texas have a similar number of Starbucks (559 and 604, respectively) and are hence the same color (and similar area) on the cartogram. However, the actual density of Starbucks is about four times higher in Washington (this is not surprising, with a Starbucks on practically every block in downtown Seattle!) due to the fact that Washington is about four times smaller than Texas. Distortions, rather than colors, are valuable in distinguishing the relative densities and countrywide distribution trends. The negative distortion for Texas (i.e. appears "too small") and the positive distortion for Washington (i.e. appears "too big") are able to visually depict the differences in coffee shop density in a way the colors cannot. This belies the value of the cartogram.

**Per capita cartograms**

The cartograms discussed throughout this paper are based upon the total values of each franchise in each state without consideration of the populations in each state. Essentially, this creates cartograms analogous a US population cartogram, except that now the
“population” is the number of Walmarts, McDonalds, or Starbucks in each state. The distortions are indicative of the density of the franchises in terms of land area.

Another interesting way of formulating the cartograms would be to divide the state franchise totals by the respective state populations and then scale the per capita values by the state areas. This creates cartograms where the distortions represent the relative state-by-state differences in franchises per capita. Positive distortions indicate states with higher than average franchises per capita, while negative distortions indicate states with lower than average franchises per capita. One way of looking at the per capita cartogram of Starbucks, for example, is how caffeinated the population is in each state due to Starbucks coffee.

Although per capita cartograms are interesting and have potential marketing applications, their analysis is outside the focus of the brief explorative study described in this paper; emphasis is instead on the spatial density of the franchises. However, just out of interest, the cartogram for Starbucks per capita is shown below.

![Cartogram for Starbucks coffee shops per capita. This can be viewed as indicating the distribution of caffeinated people across the country. The analysis of per capita cartograms for the three franchises is beyond the scope of this paper.](image)

**Summary of graphical output for franchise data**

Table 1 provides a one-page summary of the various graphical ways of displaying the distribution of Walmarts, McDonalds, and Starbucks across the nation (note the per capita cartograms are not shown here). The density-equalized cartograms are presented first, followed by three more common methods of graphical display: non-distorted maps, bar graphs, and pie charts. The plots are not meant to be readable in the small images on the summary table, but rather are intended to give a general comparison of the ways in which the data on the cartograms is often displayed.

A comparison of the graphical displays will come later in this paper. But first, I will turn to a discussion of some franchise trends and relationships that can be ascertained from the graphical output, with my main focus being the cartograms.
Table 1: Summary of Graphic Output

<table>
<thead>
<tr>
<th>Total in US</th>
<th>Max state total</th>
<th>Min state total</th>
<th>Max state area density</th>
<th>Min state area density</th>
<th>Max state per capita</th>
<th>Min state per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>1163</td>
<td>144 (CA)</td>
<td>1 (ND, SD)</td>
<td>0.51/100 sq.mi. (CT)</td>
<td>0.001/100 sq.mi. (WY, SD, AK)</td>
<td>0.14/10000 people (NH)</td>
<td>0.007/10000 people (UT)</td>
</tr>
<tr>
<td>12435</td>
<td>1165 (CA)</td>
<td>22 (ND)</td>
<td>2.9/100 sq.mi. (NJ)</td>
<td>0.004/100 sq.mi. (AK)</td>
<td>0.59/10000 people (HI)</td>
<td>0.30/10000 people (NJ)</td>
</tr>
<tr>
<td>8229</td>
<td>2010 (CA)</td>
<td>4 (VT)</td>
<td>1.7/100 sq.mi. (MD+DC)</td>
<td>0.004/100 sq.mi. (AK)</td>
<td>0.84/10000 people (WA)</td>
<td>0.05/10000 people (AR)</td>
</tr>
</tbody>
</table>

*The plots are not intended to be readable here, but are presented to give a general comparison to other ways in which the data on the cartograms is often displayed.*

![Image of cartograms](https://example.com/cartograms.png)
Franchise Trends and Relationships

Cartograms display a wealth of information, and allow the potential to study several trends and relationships. Using the cartograms created from the statewise franchise totals of Walmarts, McDonalds, and Starbucks, this section discusses a few of the more obvious and important trends—both between states and between the three franchises—that are evidenced in these cartograms.

**Walmart and McDonalds are where the people are**

The cartograms for McDonalds and Walmart are nearly identical, indicating the very similar countrywide density distributions of these two franchises. Perhaps Walmart shows a slight favor towards the Midwest and McDonalds towards the South, but the distortion differences are slight. Their distributions turn out to practically mirror the state populations, as can be seen by comparing the McDonalds and Walmart cartograms with the nearly identical population cartogram (Figure 3). Recall that it is the distortions rather than the actual state sizes that indicate the distribution trends. Positive distortions represent higher franchise densities, while negative distortions represent lower franchise densities. Hence, like the American population, McDonalds and Walmarts are found in higher densities along the west coast and the eastern half of the US, but are only scattered across the Great Plains and Rockies (i.e. the states between the west coast and Midwest).

![Cartograms](image_url)

*Figure 3. (a) Walmart cartogram. (b) McDonalds cartogram. (c) Starbucks cartogram. (d) State population cartogram. Cartograms a, b, and d are very similar, indicating that the distributions of McDonalds and Walmarts are similar to the population distribution. Starbucks has a distinctly different distribution.*
Starbucks - a west coast phenomenon

Things get interesting when we look at the cartogram for Starbucks (Figure 3). The distortions in the Starbucks cartogram show some obvious differences from the other cartograms, indicating that the country-wide distribution of Starbucks is not only different than McDonalds and Walmarts, but also that it deviates from the US population density. Larger positive distortions along the west coast imply a higher density of Starbucks in this area—namely in Washington, Oregon, California, Nevada, and Arizona. With the franchise origins dating to a 1971 debut at a small coffee shop in downtown Seattle, it should not be a surprise that Starbucks favors the west.

The distortions on the eastern half of the country are generally less noticeable, indicating generally average densities of Starbucks east of the Mississippi River, despite the higher populations in this part of the country. Lower than average densities of Starbucks are evidenced by negative distortions in many of the southern states, where it is apparently too hot for coffee. Overall, this gives rise to lower Starbucks per capita values in the east compared to the west coast. There are some notable exceptions—Virginia, Maryland, and DC in particular—where large positive distortions indicate a high per capita number of Starbucks. The government is apparently powered by Starbucks, even if a majority of the eastern US is not.

Like the other two franchises, the density of Starbucks is low in the sparsely populated Great Plains and Rocky Mountain areas. Again, there is a notable exception: Colorado has a fairly high density of Starbucks; this relationship was not seen on the other cartograms, suggesting that skiers prefer coffee over burgers and supercenters.

The most obvious state blob

A special note should be made for California, the most obvious state blob on all of the cartograms. As the most populated state (and the third largest after Alaska and Texas), California has the highest number of all three franchises. The state’s positive distortion on all of the cartograms (it is significantly bigger than either Texas or Alaska, for instance) indicates that the statewide density of each of the three franchises is higher than the national average, and that the high total is not just due to sheer size. California is definitely a very influential commercialized sector of the country.

Advantages of Cartograms

Cartograms provide a new and fun way of looking at populations (or other analogous statistics such as the number of Walmarts, McDonalds, or Starbucks). This and other advantages of cartograms have been evidenced and alluded to throughout this paper so far. This section provides a more formal discussion of some of the major advantages.

Emphasize populations and eliminate visual bias

Distortions due to density-equalization allow a cartogram to clearly depict the disparities in the populations of different regions, emphasizing weights and concentrations in a way the other graphs or color-coding cannot. A cartogram is a good way to eliminate visual bias in the visual presentation of spatially distributed data. As an illustration of these ideas, Figure 4 shows two maps of the US sequentially color-coded by the number of Walmarts in each state. The first map is the standard non-distorted view, while the second map is the
density-equalized cartogram. While the standard map gives equal emphasis to all areas, the cartogram emphasizes the Walmart-saturated smaller eastern states and deemphasizes the Walmart-deprived larger western states. In this way, the cartogram more clearly depicts the actual distribution of Walmarts than the non-distorted map.

![Figure 4](image)

**Figure 4.** (a) Typical map presentation and (b) cartogram color-coded by the number of Walmart stores in each state. The cartogram more clearly depicts the actual distribution of Walmarts than the non-distorted map.

**Better than pie charts and bar graphs**

Apart from maps, two other common ways of displaying this kind of data are pie charts and bar graphs. Again, the cartogram has advantages. There is no information in a pie chart or bar graph that cannot be shown in a cartogram—the size of the wedges or the length of the bars can be depicted via area shapes and color-schemes. The added benefit of the cartogram is displaying the spatial relationships between the data points (for example, the concentration of Walmarts in the eastern US), which cannot be achieved by pie charts and bar graphs. Moreover, pie charts and bar graphs are particularly ineffective when they become cluttered with several data points of similar values (Figure 5); this does not present a problem on a cartogram.

![Figure 5](image)

**Figure 5.** A relatively ineffective and cluttered pie chart for the number of Walmart stores in each state.

**Clear comparison of datasets**

Another advantage of the cartogram is that it allows easier visual comparison between datasets. It is much easier to compare the three cartograms in Figure 3 then to compare the same set of data in the cluttered bar graph in Figure 6.
Disadvantages of Cartograms

But the cartogram is not perfect, and there are a couple of disadvantages to this graphical method, discussed in this section.

Added complexity of creation

Constructing a cartogram usually involves using software and input files, which might not be easy to acquire or understand. For instance, the cartograms I have presented in this paper involved: finding mapping software that could create cartograms (fortunately, I didn't have to write my own program to do this), acquiring a high-resolution state outline in shapefile format to use as the map input, formatting the franchise distribution data so it could be linked to appropriate regions on the map, troubleshooting my way through running an unfamiliar program, and finally writing code so that the output files could be
presented and analyzed in R. It would have been much simpler (and quicker!) to simply load the data into Excel and create ineffective and cluttered bar graphs and pie charts.

**Confusing distortions**

Perhaps the main disadvantage of cartograms is the very distortion that defines a cartogram. If a cartogram appears too confusing and unfamiliar, people might revert to the safety of bar graphs and pie charts. High levels of distortion can lead to unrecognizable geographic features, and the map becomes ineffective. However, this is a problem that can usually be avoided by finding an appropriate level of data binning for a given cartogram. A very fine level of data binning will cause substantial local distortions, while a coarser level of binning will result in a cartogram whose features are easier to recognize, but with a less accurate impression of the true population distribution. The key is to bin the data appropriately, not too finely or coarsely, and then the cartogram achieves an ideal balance of informative distortion and recognizable geometry.

**Conclusions**

In this paper, I have presented an investigation of Gastner and Newman’s method for constructing density-equalizing cartograms. Building my investigation upon an interesting graphical analysis of the distribution of Walmarts, McDonalds, and Starbucks across the United States, I have shown that cartograms provide an invaluable tool for the unbiased visual presentation and analysis of geographic data.

There are numerous possible applications of cartograms besides determining that Starbucks has a greater following on the west coast and that golden arches and “saving money and living better” are equally distributed among the population (although this playful topic is a valuable study in itself, and something marketing agencies might in fact be interested in). Density-equalized cartograms can be used, for instance, to represent almost any human data, such as visualizations of gross national products, crime rates, election results, and so forth. There might also be potential applications outside geography, even in a three-dimensional sense. Density-equalized cartograms have a bright future.
Appendix

Data sources
The following section gives an annotated summary of the various sources of data I used.

The Idea and background
This paper investigates a technique for creating density-equalizing maps or "cartograms," as developed and described in the recent paper:


Other related references on cartograms are:

- Newman's webpage at the University of Michigan providing computer software and code for making cartograms: http://www-personal.umich.edu/~mejn/cart/.
- Website of the Worldmapper Project, where a group of researchers at the University of Michigan are gathering together an ever-growing collection of cartograms showing all sorts of aspects of the social, economic, and geographic world: http://www.worldmapper.org/.

Program to create cartograms
Dr. Frank Hardisty at the University of California in Santa Barbara used the method developed by Newman and Gastner to write some Java software that allows cartograms to be created from input shapefiles and datasets. The program outputs a shapefile corresponding to the desired density-equalized map. Software such as R can be used to view and analyze the output shapefile and associated data. Hardisty provides the cartogram software for free download on his website:


Map shapefile
The Behavioral Risk Factor Surveillance System has developed an online mapping tool. Among a wealth of datasets, they also provide a downloadable high-resolution state outline in standard shapefile format, which is what I used to create my maps.

Data

The three datasets for the state-by-state distribution of the three franchises—i.e. Walmart, McDonalds, and Starbucks—come from two different sources. The McDonalds data comes from a dataset associated with a study done by Dr. Steven Graves at California State, while the Starbucks and Walmart data come from StateMaster, an online statistical database that has compiled information from a variety of primary sources for each state. Also, the 2009 US Census data was used for the state populations.


R Code

After using Hardisty’s Javascript program to create the shapefiles for the distorted density-equalized cartograms, it was necessary to plot and otherwise analyze the data output using R. The following section provides some of my R code as reference.

Maps and plots for Walmart data (same process for McDonalds and Starbucks)

```R
library(maptools)

# create color schemes for maps using RColorBrewer
cols1 <- c("ffffff",brewer.pal(9, "Blues"))
cols2 <- c("ffffff",brewer.pal(9, "Greens"))
cols3 <- c("ffffff",brewer.pal(9, "YlOrRd"))

# plot non-distorted map from shapefile data
dev.new()
c1 <- readShapePoly('output/statemap.shp')
#names(nc1)
rrt <- nc1$WALMARTS # vector of walmarts for each state
brks <- c(quantile(rrt, seq(0,1,1/8))*0.90,max(rrt)*1.001) # define breaks for color shading
in 10% increments
spplot(nc1, 'WALMARTS', aspect = 0.65, col.regions=cols1, at=brks,
sp.layout=list(list("sp.text",getSpPPolygonsLabptSlots(nc1),
as.character(nc1$STATE_NAME),cex=0.5)),
main="Number of Walmarts by State", sub="(no distortion, color increments in 10% quantiles)"

# piechart
dev.new()
temp<-matrix(c(rrt,1:51),51,2)
temp<-temp[sort.list(temp[,1]), ] # sort in order of number of walmarts
pie(temp[,1],labels=nc1$STATE_NAME[temp[,2]], col=hcl(h=250,c=60,l=c(100:50)))
main="Number of Walmarts by State", sub="Total Walmarts in US= ", sum(rrt))

# bar graph/table
dev.new()
par(las=2) # make label text perpendicular to axis
par(mar=c(5,8,4,2)) # increase y-axis margin.
barplot(temp[,1],main="Number of Walmarts by State",
```
horiz=TRUE, col=hcl(h=250, c=60, l=c(100:50)), names.arg=nc1$STATE_NAME[temp[,2]], cex.names=0.6, sub=paste("Total Walmarts in US = ", sum(rrt)))

# plot distorted map from shapefile data
dev.new()
nc1 <- readShapePoly('output/WALMARTSTOTAL.shp')
rrt <- nc1$WALMARTS # vector of walmarts for each state
brks <- c(quantile(rrt, seq(0,1,1/8))*0.90, max(rrt)*1.001) # define breaks for color shading in 10% increments
spplot(nc1, 'WALMARTS', aspect = 0.65, col.regions=cols1, at=brks, sp.layout=list(list("sp.text",getSpPPolygonsLabptSlots(nc1), as.character(nc1$WALMARTS),cex=0.5)), main="Number of Walmarts by State", sub="(distorted to reflect number, color increments in 10% quantiles)")

# plot distorted map from shapefile data, per 10,000 people data
dev.new()
nc1 <- readShapePoly('output/WALMARTSCAPITA.shp')
rrt <- nc1$WALMARTS.
brks <- c(quantile(rrt, seq(0,1,1/8))*0.90, max(rrt)*1.001)
spplot(nc1, 'WALMARTS.', aspect = 0.65, col.regions=cols1, at=brks, sp.layout=list(list("sp.text",getSpPPolygonsLabptSlots(nc1), as.character(nc1$WALMARTS.),cex=0.5)), main="Walmarts per 10,000 people by State", sub="(distorted to reflect number, color increments in 10% quantiles)")

Barplot of all 3 franchises

nc1 <- readShapePoly('output/statemap.shp')
temp<-matrix(c(nc1$MCDONALDS,nc1$WALMARTS,nc1$STARBUCKS,1:51),51,4)
temp<-temp[sort.list(temp[,1]), ] # sort in order of #mcdonalds
counts<-as.table(temp[,1:3]) # make table for barplot

par(las=2) # make label text perpendicular to axis
par(mar=c(5,8,4,2)) # increase y-axis margin.
barplot(t(counts),horiz=TRUE,col=c("yellow","blue","green"),beside=TRUE,names.arg=nc1$STATE_NAME[temp[,4]], main="State totals of McDonalds, Walmarts, and Starbucks",cex.names=0.6)
legend(1500,50,c("Starbucks","Walmart","McDonalds"),col=c("green","blue","yellow"), pch=c(15,15,15), cex=0.8)