

PLOS Science Wednesday: Hi Reddit, we're Alasdair and Garrett and we drew a new map of the United States based on commuter data instead of traditional borders, creating new ways of interpreting how geography impacts our lives – Ask Us Anything!

PLOSScienceWednesday¹ and r/Science AMAs¹

¹Affiliation not available

April 17, 2023

[REDDIT](#)

PLOS Science Wednesday: Hi Reddit, we're Alasdair and Garrett and we drew a new map of the United States based on commuter data instead of traditional borders, creating new ways of interpreting how geography impacts our lives – Ask Us Anything!

PLOSSCIENCEWEDNESDAY [R/SCIENCE](#)

[removed]

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:
January 12, 2017

DOI:
10.15200/winn.148414.42673

ARCHIVED:
January 11, 2017

CITATION:
PLOSscienceWednesday ,
r/Science , PLOS Science
Wednesday: Hi Reddit, we're
Alasdair and Garrett and we
drew a new map of the United
States based on commuter
data instead of traditional
borders, creating new ways of
interpreting how geography
impacts our lives – Ask Us
Anything!, *The Winnower*
4:e148414.42673 , 2017 , DOI:
[10.15200/winn.148414.42673](https://doi.org/10.15200/winn.148414.42673)

© et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.

Can you give a quick link to the maps.

[RoastMeAtWork](#)

Garrett: In addition to the maps included with the paper on the *PLOS* website, we've uploaded the maps and data set to [our Figshare repository](#). Also, you'll find an alternative set of maps with zoomed-in extracts of megaregions around the US [here](#).

Your research called to my mind Chris Alexander's statements in *A Pattern Language* about the "right" size of political divisions (it's [the very first pattern in the book](#), actually). Alexander et al present four arguments, the first being...

There are natural limits to the size of groups that can govern themselves in a human way. The biologist J. B. S. Haldane has remarked on this in his paper, "On Being the Right Size":

... just as there is a best size for every animal, so the same is true for every human institution. In the Greek type of democracy all the citizens could listen to a series of orators and vote directly on questions of legislation. Hence their philosophers held that a small city was the largest possible democratic state.... (J. B. S Haldane, "On Being the Right Size," *The World of Mathematics*, Vol. II, J. R. Newman, ed. New York: Simon and Schuster, 1956, pp. 962-67).

It is not hard to see why the government of a region becomes less and less manageable with size. In a population of N persons, there are of the order of N^2 person-to-person links needed to keep channels of communication open. Naturally, when N goes beyond a certain limit, the channels of communication needed for democracy and justice and information are simply too clogged, and too complex; bureaucracy overwhelms human processes.

And, of course, as N grows the number of levels in the hierarchy of government increases too. In small countries like Denmark there are so few levels, that any private citizen can have access to the Minister



of Education. But this kind of direct access is quite impossible in larger countries like England or the United States.

We believe the limits are reached when the population of a region reaches some **2 to 10 million**. Beyond this size, people become remote from the large-scale processes of government. Our estimate may seem extraordinary in the light of modern history: the nation-states have grown mightily and their governments hold power over tens of millions, sometimes hundreds of millions, of people. But these huge powers cannot claim to have a natural size. They cannot claim to have struck the balance between the needs of towns and communities, and the needs of the world community as a whole. Indeed, their tendency has been to override local needs and repress local culture, and at the same time aggrandize themselves to the point where they are out of reach, their power barely conceivable to the average citizen.

I count 53 communities in your map, which averages some **6 million people** in each. Is this just a big coincidence, or do you think we're seeing some emergent organization of social networks toward a 'natural' size optimum?

[TootZoot](#)

Garrett: This is a very important question which, to be answered properly, requires invoking political theory, sociology, and the philosophy of organismic complexity! I can't do all of those topics justice here, but both the Haldane and Alexander pieces you quote above are good examples of the theory that there is some "natural" size of human community, above which size organized social functions tends toward chaos. This is a very old idea: Aristotle, for instance, also argued that there was a correct size for the ideal community, summed up by the rather tricky concept of "self-sufficiency."

In our case, we artificially limited the community-detection software to produce 50 communities, a number based on the current count of US states (the final map, as noted, includes some subjective toying with the purely computational boundaries). However, these are *not* limited by equal population. Some regions are an order of magnitude larger than others.

Here again, we face the tricky question of "self-sufficiency." Imagine two small communities that *only* have economic interactions with each other. Then imagine a massive conurbations of cities deeply interlinked. Does the former, isolated region deserve to be "one" place, and the latter, interlinked region also deserve to be just "one" place? Herein lies the paradox at the center of this research!

Hi! Thanks for doing this AMA.

My question is: why is this important? What do you think the largest impact of this result should be?

To unpack that question a bit more, Do you think this should influence public policy with regards to infrastructure investment? How about political policy? Looking at some of your divisions, it would be hard to color the map into "red" and "blue" so I'm not sure how much political ideology factored into your algorithms. It seems to have more of an urban vs rural split.

[CallMeDrewvy](#)

Garrett: Probably the biggest takeaway from this research is that borders inherited from centuries ago are not always the most useful geographic categories for understanding the functional relationships of the present day. But those old borders, of course, still have power: they control where we vote, where we pay taxes, where laws are enforced, and so on. Our hope is that by offering an evidence-based analysis of how people move around, we can push both administrators (like transport planners) as well as ordinary citizens rethink what geographic categories are most relevant today.

Very cool! I'm a GIS Analyst and have always loved unique maps, and new methods of cartography. Simple question... What software are you using to collect and process the data?

[DavidAg02](#)

Alasdair: thanks David. We used QGIS for the mapping, and something called Combo for processing the data and dividing up the commutes into regions. Combo was developed at MIT - http://senseable.mit.edu/community_detection/. The only problem we had is that the 4 million commuter lines clogged up our computers so we had to use cloud computing to crunch the data - for that we used Amazon Web Services.

Gerrymandering political borders is a big issue currently in the US. How might a government adopt a system where district borders are drawn by a computer algorithm instead of the ruling party? Has this been done already in some parts of the world, and if so, how's it working?

[llamachemist](#)

Alasdair: that's a great question. Neil Freeman took our boundaries and used it in a really interesting piece of analysis on this subject. Here's an extract on it from the Washington Post: https://www.washingtonpost.com/news/wonk/wp/2016/12/19/five-ways-of-redrawing-the-map-that-would-change-the-2016-election/?utm_term=.65aa751d5997

I'm intrigued by state borders that seem to act as psychological barriers. I'm in Wisconsin. That southern state border is a mental obstacle to people in Wisconsin. I've met people who've never left the state. People will travel between Madison and Milwaukee to visit an event or a new restaurant, but the similar distance to Rockford, IL seems "too far."

[jfoust2](#)

Alasdair: I really love this comment, because it highlights the importance of things that we sometimes can't quantify or measure easily. Sometimes distance is the only thing that matters, but more often there are other factors too. The kinds of mental barriers that exist mean travel patterns are not always what we might expect if we only think about distance.

I'm intrigued by state borders that seem to act as psychological barriers. I'm in Wisconsin. That southern state border is a mental obstacle to people in Wisconsin. I've met people who've never left the state. People will travel between Madison and Milwaukee to visit an event or a new restaurant, but the similar distance to Rockford, IL seems "too far."

[jfoust2](#)

Garrett: I was also surprised at the line that almost perfectly traces the WI-IL border, especially given that these two states have a tax reciprocity agreement. It just goes to show how old boundaries have the ability to reproduce themselves. Folks from Janesville might think of people from Wausau as being more their "neighbors" compared to those from Rockford—even though from a purely "natural" geography that makes no sense. How we imagine ourselves in places has huge effects on where we choose to live, travel, and who we interact with!

Hello,

I recently just graduated with a degree in Geography and GIS. I was wondering how you've seen the field of geography change as technology has gotten increasingly better. Also how did technology (programs, coding) help you sort data and create your maps?

[TwistedAnomaly](#)

Garrett: The ability to work with big data sets is definitely a crucial skill for geographic research these days! Without powerful computers (we used Amazon servers), we could never have processed the enormous amount of data included in this set. These tools make it possible to "see" evidence in data which would not have been possible before.

However, the most important skill remains the ability to ask good questions and interpret results. Computational sophistication will never fully replace analytical inquiry and real-world understanding.

First of all, interesting work you've done, the visualizations look great! I live in Sweden where there's a debate right now about what our new regions should look like. Do you happen to know if similar studies have been done here, or anywhere else in Europe for that matter!

[k3nterin](#)

Garrett: Regionalization studies have been done all across the world for more than a century using different types of evidence and geographic-analysis tools to make the case for different kinds of regional delineation. Only recently, however, have we had the kind of data-gathering and data-crunching tools which make studies like these possible.

The MIT folks who created this partitioning algorithm [used it to partition the UK](#) using telephone data. Another paper [also looks at several other European countries](#).

This is fascinating work especially since I work with diasporas. I have two questions and I apologize that they are a bit wordy.

First, I'm curious about applying a similar kind of method to national borders. For example, if you expanded the boundaries of your map beyond the US to include Canada and Mexico, what would this look like? I recognize that incorporating illegal entry could make this work complicated (and politicized). But thinking about geography, citizenship, and belonging and the ways socio-cultural borders of belonging can be porous when political borders are not this kind of mapping is really interesting. Have you tried applying this to international borders?

Second, I'd love to see utilization of this kind of mapping for ethnic borders and boundaries. Since ethnic groupings and borders do not depend upon political ones, it would be interesting to see how self identified ethnic communities are mapped onto contemporary American political boundaries and how they migrate/commute/return. But ideally this would be using data that is better than the Census, which is obviously limited and problematic for identifying ethnic communities especially sub-groups within larger categories. Are there datasets that would allow you to do that kind of mapping?

[firedrops](#)

Garrett: This is a great question, and it points to some issues with how we treat "big data" as a kind of evidence. In the case of our study, the commuter counts come from the US Census Bureau, and therefore the data set doesn't include international commutes. My hunch is that if such data were available, you'd likely see some international regions—perhaps Seattle-Vancouver, Buffalo-Niagara, or, as CptnStarkos suggests below, San Diego-Tijuana.

Because the available sources for large data sets like these tend to come from national agencies, they

have the effect of reproducing the territorial integrity of the nation itself. That's one of the reasons why we should always be careful about claiming that data analysis leads to some perfectly "correct" or "true" solution.

Regarding ethnic groupings: the Cooper Center's [racial dot map of the United States](#) is an absolute must-see on this topic. While it doesn't include commutes (just home locations, aggregated by block group), it still makes residential segregation shockingly apparent.

If we could find good data about individuals race/ethnic status as well as their commuter habits, it's possible you'd begin to see overlapping regions, where different kinds of people inhabit very different territories. For instance, it's likely that many of the people commuting into New York City from Westchester County are white, while a self-contained commuting group in Harlem and the Bronx could emerge. Those are just guesses, of course, but they suggest some of the ways in which a totalizing geography has its limitations.

Will you publish your code on github or similar?

What data sources are you using?

How hard would be to get more detailed maps of smaller regions. As European I am amazed by usa city borders and very very big metro areas. Eg. New York or from smaller ones New heaven North of New York. American population figures for cities are sometimes very different from metro areas.

Do you know why Metro area population is getting less focus than city?

[rzet](#)

Alasdair: you can see the data and files on Figshare here:

https://figshare.com/articles/United_States_Commuters_and_Megaregions_data_for_GIS/4110156. As for metro vs city, this is something we were trying to explore. Our 'megaregions' attempt to pick out how big some metro areas are in relation to their wider commutershed areas. As you can see, they are often pretty big.

Hi guys, Thanks for the AMA.

Were there any parts of the final map that surprised you? Were there any parts that could have gone a slightly different way? Also, Garrett, do you go by Dash instead?

[Jupes Star](#)

Alasdair: what surprised me was in some areas where our computational approach picked out state boundaries, even though the algorithm we used was not based on geography. It seems that some state boundaries also serve as a kind of barrier to travel, for a variety of reasons - as other people have suggested below. You can see more examples if you look at the full set of maps online:

https://figshare.com/articles/United_States_Commuters_and_Megaregions_data_for_GIS/4110156

I'm impressed. I've lived in 7 of those zones and traveled around them and through many more. Of where I've been, your boundaries do match up with where I felt the "city of most influence" (my internalized name) changed. It's neat to see perceptions and intuitions validated by some data mining.

[YellowBeaverFever](#)

Garrett: Yes, we were also surprised with how well a "dumb" algorithm, which doesn't know anything

about how we intuitively imagine our regional geography, was able to produce such a recognizable map.

Some of your commuting regions seemed arbitrarily large, incorporating several major commuting central cities in large regions whereas in other areas you have small distinct commuting zones. Examples of both.

a) Arbitrarily large - in the [/r/DataIsBeautiful](#) rendering of your map, the area labeled Central Florida on the map linked in this post was labeled "Tampa". This appeared to incorporate extended commuting for Tampa/St Pete, Orlando, Florida Space Coast, Ocala/Gainesville, and Lake City into a single region. There are very definitely multiple central draws in this mega region with Tampa/St Pete and Orlando being major centers. Regional centers like Ocala, Gainesville, and Lake City would all be fairly small, and one could argue combining them into a single entity. However, I have to question why you would combine all of these areas into a mega region. Clearly no one is commuting from the far edges of this region to the major centers -- e.g. nobody commutes from Lake City to Orlando or Tampa/St Pete. Similarly, nobody commutes from the Space Coast to Lake City or Gainesville. What was your criteria for the aggregation instead segregation into multiple overlapping centers?

A similar argument can be made for other very large regions like the New Orleans and Delta Region. Florida Panhandle commuting can be drawn to Mobile, Pensacola or military complexes further east at Fort Walton/Panama City. New Orleans has its own draw, spilling into both coastal Louisiana and Mississippi. Jackson and Memphis are each regional commuting centers

b) Multiple Distinct - your linked map has this combined in the Columbia plateau region, but the colored map on data is beautiful actually had multiple small regions in eastern Washington state around Spokane, Yakima and the Tri-Cities Richland/Kennewick and Walla Walla. Why were they separated out when the megaregions weren't? Why have you now aggregated them?

[shiningPate](#)

Garrett: What's important to note about the size of the regions is that they *don't* mean that everybody is commuting to a single place. In fact, more people commute long distances to get to NYC than they do to New Orleans, but New Orleans is set inside a much larger region than NYC is in this research.

The reason is that we're trying to find *collections of relatively interdependent nodes*. For instance, if point A is strongly connected to B, and B strongly to C, and C to D, then those four points seem to make up a whole region even if very few people are commuting from A to D.

Meanwhile, you can *always* drill down smaller, detecting sub-communities within sub-communities even down to the neighborhood level. So the problem is to figure out where there are meaningful breaks in the pattern.

In reply to your follow-up: the data makes no distinction about mode of transport. All we know is how many people leave one census tract (as their home, or origin tract) and arrive at another census tract (as their work, or destination tract).

Some of your commuting regions seemed arbitrarily large, incorporating several major commuting central cities in large regions whereas in other areas you have small distinct commuting zones. Examples of both.

a) Arbitrarily large - in the [/r/DataIsBeautiful](#) rendering of your map, the area labeled Central Florida on the map linked in this post was labeled "Tampa". This appeared to incorporate extended commuting for Tampa/St Pete, Orlando, Florida Space Coast, Ocala/Gainesville, and Lake City into a single region.

There are very definitely multiple central draws in this mega region with Tampa/St Pete and Orlando being major centers. Regional centers like Ocala, Gainesville, and Lake City would all be fairly small, and one could argue combining them into a single entity. However, I have to question why you would combine all of these areas into a mega region. Clearly no one is commuting from the far edges of this region to the major centers -- e.g. nobody commutes from Lake City to Orlando or Tampa/St Pete. Similarly, nobody commutes from the Space Coast to Lake City or Gainesville. What was your criteria for the aggregation instead segregation into multiple overlapping centers?

A similar argument can be made for other very large regions like the New Orleans and Delta Region. Florida Panhandle commuting can be drawn to Mobile, Pensacola or military complexes further east at Fort Walton/Panama City. New Orleans has its own draw, spilling into both coastal Louisiana and Mississippi. Jackson and Memphis are each regional commuting centers

b) Multiple Distinct - your linked map has this combined in the Columbia plateau region, but the colored map on data is beautiful actually had multiple small regions in eastern Washington state around Spokane, Yakima and the Tri-Cities Richland/Kennewick and Walla Walla. Why were they separated out when the megaregions weren't? Why have you now aggregated them?

[shiningPate](#)

Alasdair: you're right in relation to the question of long commutes. However, you would be very surprised by some of the extreme commutes, which you can see on p.8-10 of this if you're really keen: http://eprints.whiterose.ac.uk/89361/7/WRRO_89361.pdf The point is that the 'megaregions' are a kind of geographic container for smaller sub-regions and your FL example is a good one. But, within these wider areas there are some very long 'mega-commutes', for sure. On b), the final map involved a bit of human interpretation as well. It's the only one where we tried to combine our decision making with that of the algorithm. Think of it as a map for discussion and debate rather than a final product! Thanks for these comments.

Does your research give you a way to rank the regions you have identified by economic self-reliance? Which regions have the best integration along verticals of economic activity?

[chodpaba](#)

Alasdair: no, but this could be a really useful next step for us. Our first task was to use a computational approach to see what the algorithm we used would produce - and assess how valid the results were.

I have a question that relates to arguments in support of the logic used to defend the electoral college.

I am a native San Franciscan and from college onwards I've constantly heard remarks of how rare it is to meet someone who grew up here (or the Bay Area, and personally I have rarely met people who moved from other parts of California).

If we are saying some states need their votes to be counted more to represent the people in them, that would mean the migration patterns would have to be such that people do not relocate to other states in significant numbers. From my own observational experience and some recent explorations through census data, as well as the ease of transportation which the writers of the constitution could not have anticipated, I am curious if the reality of migration could invalidate that argument.

Have you spent any time studying migration patterns or have any insights or hunches on this from your studies so far?

[Youreagoomba](#)

Alasdair: the short answer is no we haven't. But, your point is very important because it highlights another issue we raise in our paper - that commuting is only one kind of connection between places. There are many more that matter. Migration - particularly for the Bay Area - is obviously going to have a big impact.

Hey! The History class I was just in talked about your maps! So my question is, what made you come up with the maps?

[James Westen](#)

Alasdair: the basic answer is curiosity. We were interested to see what would happen if we tried to divide the country up based on patterns of commuting, but using a computational approach rather than a cultural or historical one. As you can see, the results do make some sense geographically, but of course we realise they are not perfect! In fact, our work was in part a test to see whether you can use algorithms on their own to do this kind of thing. Our conclusion is that you also need human input. We use commuting because it is a very important part of the economic functioning of the nation.

I remember seeing a similar project a few years ago that asked people what city they identified with and drew borders around the cities based on the results.

I can't remember the name of it, but I was wondering if you were aware of it and had any comments on how similarities/differences with this one. Obviously this is divided by states and not cities but other than that I remember looking largely similar, so that's interesting.

[shadow1515](#)

Garrett: Geographers have long been interested in the concept of "mental mapping" to get a sense of where people imagine the centers and edges of their own places. Recently, interactive web mapping has opened up new possibilities as well. Here's a project asking people to [draw their neighborhoods in Boston](#); Alasdair used the same technique to study [metro areas in the UK](#), and I used it to ask people to [define the NH-VT "Upper Valley"](#) where I now live.

I remember seeing a similar project a few years ago that asked people what city they identified with and drew borders around the cities based on the results.

I can't remember the name of it, but I was wondering if you were aware of it and had any comments on how similarities/differences with this one. Obviously this is divided by states and not cities but other than that I remember looking largely similar, so that's interesting.

[shadow1515](#)

Alasdair: there have been quite a few of these kinds of studies in the past, including excellent work in Boston by Bostonography: <http://bostonography.com/2013/neighborhoods-as-seen-by-the-people/> I also did one of these for cities and people drew cities across the world: <http://ajrae.staff.shef.ac.uk/nhood/#view> The question of individual identities and regions is really a very important one. That's why we think it's so important not to trust empirical analysis alone or to assume a kind of factual accuracy or truth in it. It's also a good way to start a debate on the topic.

Thank you for doing this AMA, very interesting research.

How would you compare and contrast these regions vs. a map split by metro/micropolitan statistical areas?

Can you explain the reasoning for limiting the number of regions to 50, other than it correlates roughly to the current number of states (48, since we're only looking at the contiguous US)? I assume it has to do with the output from Combo, but it seems like large rural areas end up getting clumped together.

One of the criticisms I read in response to your paper was that it tended to group regions that might have very different economic bases but were grouped together based on relative proximity as well as enough commuter traffic between various nodes in the region to consider them similar. Any thoughts on how to better differentiate between "people drive between these places" and "these places are linked because they create and consume products within this same region"?

Any further research planned for this topic? Any thoughts on using input-output economic models to create regions?

Thank you for your time.

[_rlease](#)

Garrett: Proximity actually isn't included at all in the algorithmic method (see [here](#) for an example illustration). We only look at the strength of commuter relationships. But you're right: commutes are only one proxy for economic integration (though we argue that they are a very good one, especially in an increasingly service-oriented economy). We could alternatively take different measures of connection, like package shipments, and see if that results in similar or different regionalization results.

Thank you for doing this AMA, very interesting research.

How would you compare and contrast these regions vs. a map split by metro/micropolitan statistical areas?

Can you explain the reasoning for limiting the number of regions to 50, other than it correlates roughly to the current number of states (48, since we're only looking at the contiguous US)? I assume it has to do with the output from Combo, but it seems like large rural areas end up getting clumped together.

One of the criticisms I read in response to your paper was that it tended to group regions that might have very different economic bases but were grouped together based on relative proximity as well as enough commuter traffic between various nodes in the region to consider them similar. Any thoughts on how to better differentiate between "people drive between these places" and "these places are linked because they create and consume products within this same region"?

Any further research planned for this topic? Any thoughts on using input-output economic models to create regions?

Thank you for your time.

[_rlease](#)

Alasdair: one of the issues with any computational approach or algorithm is that the user must make some decisions on parameters, etc. We ran the analysis several times and it just happens that 50 produced a result that 'made sense' from our perspective in relation to what we knew about the economic geography of the US. But, yes, this is where imperfections come in. The old statement that 'all models are wrong, but some are useful' (George Box) is also true here. We think this is quite useful, but it's far from perfect when you look closely. We're going to explore this kind of work further - probably I will look at the UK next.

