# Using human geography to build a more meaningful compactness measure for districting algorithms

Zhenghong Lieu

14 November 2019

**Abstract**

[This is a working title, and everything about this is very much a work-in-progress.]

Most existing districting approaches aim to optimise over geographical compactness measures like Polsby-Popper and Convex Hull. Compactness is important because compact districts better represent communities of interest and have been shown to improve democratic effectiveness. However, compactness measures are imperfect proxies for human interaction as they cannot account for human geography. For instance, a typical compactness measure would put two villages separated by a big mountain together, even if these villages have near-to-zero interaction. While the shortcomings of compactness measures are well-known in the literature, existing approaches to improve it have failed due to lack of data and computational intractability (NP-hardness). I develop a new metric ("human compactness") based on travel times that is computationally feasible—yet still captures human geography—and calculate it for many US Congressional districts. Finally, I augment several districting algorithms with the metric, including the Metric Geometry and Gerrymandering Group (MGGG's) Monte Carlo Markov Chain (MCMC) districting algorithm, and show that the plans drawn with my metric are superior in electoral competitiveness and media congruence, which directly translates to increased federal funding.

# Introduction

Most existing districting approaches aim to optimise over geographical compactness measures like Polsby-Popper and Convex Hull. Compactness measures are important because they encode a notion of geographic proximity. A wealth of empirical evidence suggests that democratic effectiveness decreases when districts are noncompact, due to the lack of shared communities of interest and geographic proximity.

However, compactness measures are imperfect proxies for human interaction as they cannot account for political geography. For instance, a typical compactness measure would put two villages separated by a mountain or large body of water in the same district, even though these villages may have near-to-zero interaction.

The shortcomings of geographic compactness measures have been well-discussed in the literature, and several solutions have been proposed. Some have tried to capture human interaction directly, but such an approach is infeasible for most states as the data required does not exist. Yet others have tried to develop better measures of compactness, but their approaches were fundamentally computationally intractable. Using computer science and software engineering techniques, I develop a metric that is feasible to compute and does not require extra data, yet still captures the key features of human interaction and communities of interest. The metric, which I call "human compactness", is defined as the ratio between the sum of travel times of a voter's nearest neighbours, and the sum of travel times of his co-districtors. I show that it more robustly captures the interactions and shared interests of proximate voters.

Finally, I deploy the metric in real life by augmenting a Monte Carlo Markov Chain (MCMC) districting algorithm to use the human compactness metric.

[Ideally, I'd also need to show that the plans drawn optimising for my new metric are significantly different, but I've not done so yet and thus can't make any claims at the moment.]

## Why compactness matters

37 states require their legislative districts be reasonably compact, and 18 states require congressional districts to be compact as well (Levitt 2019). But why do we care about district compactness? There are two main reasons strongly rooted in the political science literature: representation (Pitkin) and competition (Schumpeter), both fundamental normative reasons for …

Mandating compactness increases both representation and competition.

### Representation (Pitkin)

Compactness increases representation through two mechanisms: by preventing gerrymandering, and putting people who resemble each other together.

In The Concept of Representation (1967), Pitkin identifies four views of representation.

1. Formalistic: authorisation and accountability
2. Symbolic: representative is "accepted by" the people: relates to turnout, voting knowledge, etc.
3. Descriptive: representative as a "mirror of the people": looks like, shares common interests with, sharing experiences with voters. black candidate with black voters etc
4. Substantive: representative acts on behalf of, and advances the policy preferences of the represented

All four types of representation are improved by compactness.

#### How compactness prevents harms to formalistic representation

Formalistic representation involves both authorisation and accountability. Authorisation means that the representative must have come to power through a legitimate mechanism, and accountability means that constituents must be able to punish their representatives and vote them out of office if they do a bad job.

A fair electoral system delivers both authorisation and accountability. But gerrymandering harms both of these. Gerrymanders who "pack" districts ensure safe seats for incumbents, meaning that that representatives can do a poor job and yet be assured of a large margin of victory. By definition, if representatives can stay in power regardless, they are unaccountable. More generally, if voters cannot materially affect the outcome of elections due to gerrymandering, this casts doubt on the legitimacy—and thus authorisation—of the representatives.

In order to gerrymander effectively, however, gerrymanderers must carve out districts in profoundly unnatural ways. This means that gerrymandered districts are often very noncompact. As Polsby and Popper (1991) put: "Without the ability to distend district lines ... it is not possible to gerrymander. The diagnostic mark of the gerrymander is the noncompact district." Mandating compactness prevents gerrymandering, and thus prevents representational harms. The courts have long relied on compactness to identify and prevent gerrymandering. In *Davis v. Bandemer*, Justices Powell and Stephens pointed to compactness as a major determinant of partisan gerrymandering... [expand on this: how?]

### How compactness improves symbolic representation

Compactness also improves symbolic representation, which measures the degree to which constituents support and accept their representative. Arzheimer and Evans (2012) find that constituents support more strongly candidates that live close to them, even controlling for strong predictors of vote choice like party feeling and socio-economic distance. Because districts in which people generally live near each other are more compact than one in which they do not, compact districts are more likely to have candidates and voters close together.

Media knowledge: if voters don't even know who their candidate is, obviously they can't really have "accepted" them. Cite Snyder and Stromberg here.

While turnout is affected by many factors apart from support for one's representative, voter roll-off...

Stephanopoulos finds that the rate of voter roll-off is significantly higher in spatially diverse districts. This is significant because all else being equal, less compact districts will usually be more spatially diverse: their odd shape and long reach means that they will include radically different neighbourhoods.

### How compactness prevents harms to descriptive representation

Descriptive representation involves the extent to which a representative "mirrors" his constituents: this could be belonging to the same race or socioeconomic class, sharing common experiences, or being part of the same communities of interest.

In order for a representative to mirror his constituents, his constituents must be somewhat homogeneous. A single representative cannot resemble multiple highly heterogeneous populations at once. Professor Bruce Cain writes that if a district is spatially "divided between nonwhite and white, rich and poor, rural and urban,", "then it may be very hard for one representative to represent all factions well." And Professor Thomas Brunell contends that "the more homogeneous a district, the better able the elected official is to accurately reflect the views of more of his constituents."

But gerrymandered districts frequently contain highly heterogeneous populations. Take the example of racial gerrymandering: in order to build majority-minority districts as mandated by law, yet maintain their political hegemony, gerrymanderers will draw highly noncompact districts that put together completely heterogeneous populations with only one thing in common: their race. Descriptive representation suffers as a result.

Indeed, the Supreme Court has previously found such gerrymandered districts unconstitutional. In the 1993 case of Shaw v. Reno (Shaw I), the court struck down a North Carolina district which slithered "in snakelike fashion through tobacco country, financial centers, and manufacturing areas, gobbl[ing]... enclaves of black neighborhoods" in each of these disparate regions. And in the 1995 case of Miller v. Johnson, the Court struck down a Georgia district because the two populations it connected were "260 miles apart in distance and worlds apart in culture" (Stephanopoulos 2011).

Once again, a compactness requirement prevents gerrymandering and thus prevents harm to descriptive representation.

### How compactness improves substantive representation

Substantive representation is about the representative acting on behalf of, and advancing the interests of, his constituents.

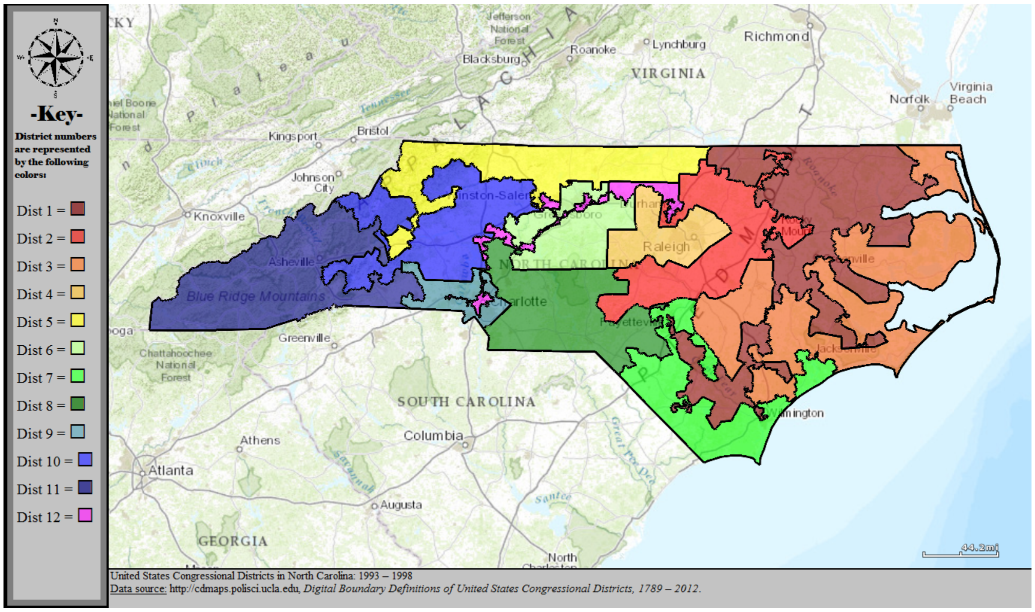


Figure 1: North Carolina's 1993 congressional districts: offending district in pink

Stephanopoulos pg. 1920:

> In two well-known series of interviews carried out by political scientists, elected officials in Congress 61 and state legislatures 62 repeatedly stated that they found it difficult to represent spatially diverse districts. House members complained that they could not easily discern the "lowest common denominator of interests" in geographically varied districts, 63 while state legislators expressed frustration that they "sim- ply [could not] 'represent' the views of . . . diverse groups when there are sharp conflicts." 64 More conventional studies confirm that repre- sentation (in the sense of responsiveness to constituent interests) is in- versely related to districts' top-line demographic, economic, and ideo- logical diversity.

see Snyder and Stromberg: compactness increases congruence which increases media knowledge, which means more voter oversight, which means more federal funding.

See also game-theoretic analysis: less compact districts means less efficient public good provision

## Competitiveness (Schumpeter)

Increased media congruence increases challenger vote share

Compactness -> congruence -> challenger vote share

> The most common defense of spatially diverse districts, of course, is that they are necessary to foster electoral competition. Competition is essential to a properly functioning democracy, the argument goes, and heterogeneous districts are more likely to be competitive than homogeneous districts. Fortunately, we do not actually have to choose between homogeneity and competition. It turns out, empirically, that the two characteristics are weakly correlated, if at all, **and that it might be more geographically uniform districts that in fact are more competitive**. In particular, several studies have found that competitiveness in both general and primary elections is unrelated to constituencies' top-line demographic, economic, and ideological diversity, and that quality challengers are no more likely to materialize in heterogeneous districts than in homogeneous districts. ( p. 1922 – 1923)

He claims that challengers' vote shares are typically higher when districts better correspond to geographic entities like political subdivision and media markets (James Campbell et al.)

Seat-vote curve by Gary King: gerrymandered districts

# Why compactness matters (OBSOLETE)

37 states require their legislative districts be reasonably compact, and 18 states require congressional districts to be compact as well (Levitt 2019). But why do we care about district compactness? There are two main reasons: first, putting people who live near each other in the same electoral district increases democratic effectiveness; second, demanding compactness is a procedural safeguard against, and statistical test for, gerrymandering.

First and foremost, compactness encodes a notion of geographic distance, and there are strong normative and empirical reasons to put people who live near each other in the same electoral district. People who live near each other often belong to similar *communities of interest*: that is, they often share socioeconomic backgrounds, cultures, and interests. This makes it likely that they will share the same political goals. For instance, people living in the same neighbourhood have an interest in their local hospital, or in securing federal funding for their neighbourhood's homeless. These common political goals make it likely that they will form meaningful units of political contestation. In a review article, Campbell (2007) wrote that "Beneficiaries who are geographically near to others and who can identify [other beneficiaries]... can more easily exchange information and band together for political action".

In contrast, it's difficult for voters in a district to mobilise with others who are far away, or who (due to factors like race or socioeconomic class) do not share the same interests as them. The Supreme Court has struck down districts for exactly this reason.

In the 1993 case of Shaw v. Reno (Shaw I), the court struck down a North Carolina district which slithered "in snakelike fashion through tobacco country, financial centers, and manufacturing areas, gobbl[ing]... enclaves of black neighborhoods" in each of these disparate regions. And in the 1995 case of Miller v. Johnson, the Court struck down a Georgia district because the two populations it connected were "260 miles apart in distance and worlds apart in culture" (Stephanopoulos 2011).

A wealth of empirical evidence shows that more compact districts lead to better democratic outcomes. Stephanopolous (2011) shows that more "spatially diverse" districts have less engaged voters and poorer service provision. A spatially diverse district—with respect to income, for instance—is a district where rich and poor live in distinct regions, as compared to a spatially homogenous district where rich and poor are uniformly distributed throughout. He finds that the rate of voter roll-off is significantly higher in spatially diverse districts. This is significant because all else being equal, less compact districts will usually be more spatially diverse: their odd shape and long reach means that they will include radically different neighbourhoods. While it's possible in theory to draw a very uncompact district that is spatially homogenous (by lassoing all the poor into one district, for instance), this is rarely the case in practice. Snyder and Stromberg (2010) show that regions with greater "congruence" between newspaper markets and congressional districts have better political literacy (measured by knowledge about one's representative) and political outcomes (as measured in federal spending per capita). It seems reasonable to assume that local newspapers make decisions on where to expand their coverage based on geographic boundaries and communities of interest. For instance, a local newspaper might not want to deliver papers to readers 300 miles away from their presses due to logistical issues, and readers might in any case not be interested in faraway local news. If this is the case, then compact districts are better as they more closely track the natural patterns of newspaper market distribution, and thus directly lead to better democratic outcomes.

Finally, noncompactness has also been linked to poorer political participation. Arzheimer and Evans (2012) find that constituents support less strongly candidates that live far from them, even controlling for strong predictors of vote choice like party feeling and socio-economic distance. In part, voters strongly support proximate candidates because they think that these candidates better represent their interests. Because districts in which people generally live near each other are more compact than one in which they do not, compact districts are more likely to have candidates and voters close together. Similarly, Dyck and Gimpel (2005) find that voters living further away from a voting site are less likely to turn out to vote. *Ceteris paribus*, noncompact districts will have more voters further away from a voting site, as each voting site will serve fewer voters.

Normative reasons aside, compactness also acts as a statistical test to detect gerrymandering. Gerrymandering is performed by packing and cracking, which involves either pulling disparate blocs from far-flung communities into a singular tendril-like district, or splitting natural blocs of voters into multiple districts to dilute their influence. In order to do so, districts must be drawn in a very contorted way which a compactness test picks up on.

As Polsby and Popper (1991) put: "Without the ability to distend district lines ... it is not possible to gerrymander. The diagnostic mark of the gerrymander is the noncompact district." The courts have long relied on compactness as a marker of gerrymandering. In *Davis v. Bandemer*, Justices Powell and Stephens pointed to compactness as a major determinant of partisan gerrymandering.

## The case for travel times over geographic compactness

Because geographic compactness is so useful as a notion of geographic distance and to detect partisan gerrymandering, almost all automated districting algorithms today take into account—or optimise directly for—geographic compactness. For instance, Magleby and Mosesson (2018) take care to state that their districting plans are "relatively compact". Deford et al (2019) explicitly use geographic compactness as a score function in their Monte Carlo Markov Chain (MCMC) algorithm, and Levin and Friedler (2019) evaluate their districting plans on three different geographic compactness scores. However, I believe that automated districting algorithms should optimise for travel times instead.

Optimising over geographic compactness preserves communities of interest and thus increases democratic effectiveness. But using travel times is a strict improvement over geographic compactness. Because it is a purely geometric measure, geographic compactness neglects human geography—yet geography shapes human interaction to a large degree. Fryer and Holden Jr. (2011) wrote:

> Suppose there is a city on a hill. On the West side is mild, long incline

> toward the rest of the city, which is in a plane. On the East side is a steep cliff, either impassable or with just a narrow, winding road that very few people use. While the next residential center to the East is much closer to the hilltop on a horizontal plane, it is much further on all sorts of distances that we think might matter: transportation time, intensity of social interactions, sets of shared local public goods and common interests, etc. Thus, for all practical purposes, one probably wants to include the hilltop in a Western district rather than an Eastern one. More general notions of distance can handle this.

As they point out, it's almost obvious that communities form around travel times. People form communities with others who go to the same schools, use the same hospital, and frequent the same shops—this is a function of travel times, and not geographic compactness. As Levitt (2019) states:

> people with common interests don't generally look to geometric shapes – or even strict political lines – when they consider where they want to live.

Similarly, newspaper markets obviously don't track geometric shapes when determining expansion: instead, they consider constraints like transportation routes and community interest.

Geographic compactness may still be very useful in testing for partisan gerrymandering. But automated districting algorithms are not merely useful to test for gerrymandering: we can use them to propose new, unbiased plans. And in this respect automated districting algorithms can be shown to be unbiased without appealing to geographic compactness. We can inspect the source code of these algorithms and verify that they do not take into account a region's voting history, or try to give one party an unfair advantage. Yet, as these algorithms are random in nature, one might worry that even a "colour-blind" approach to drawing districts could produce a biased sample of the set of all possible districts. But Magleby and Mosesson (2018) show that unbiasedness can be proven by subjecting the algorithm to a battery of tests, or comparing its outputs to another known unbiased distribution of districting plans. As a procedural safeguard, therefore, geographic compactness is no longer relevant in automated districting algorithms.

## Why don't we measure communities of interest directly?

An obvious rejoinder is: if geographic compactness is only desired insofar as it encodes communities of interest, might we not just skip the middle-man and optimise for keeping communities of interest together directly?

Firstly, it is notoriously difficult to measure communities of interest. California made districting maps that respected "communities of interest" through a year-long, drawn-out process, which involved recruiting unbiased candidates to form the committee, recruiting technical experts to draw district boundaries, holding dozens of public input hearings, reading through comments and suggestions from over 20,000 individuals and groups, and conducting hundreds of field interviews. It employed a "district-by-district approach in which [it] deliberated over the best approach to minimize the splitting of cities, counties, neighbourhoods, and local communities of interest." (pg. 24)

Such a process is surely impossible for automated districting algorithms. In fact, the Californian Redistricting Commission explicitly ruled this out in an FAQ on their website:

> Q: Can't this whole process be done by a computer program?
>
> A: The Commission is relying on the active participation of citizens across California to weigh in on how the districts should be drawn, since information about "communities of interest" is not collected in the Census. This is an open conversation that will assist the Commission in evaluating citizen input and exercising responsible judgment about what districts should look like – a computer could never do that.

A citizen-driven approach to districting like the one California adopts is very admirable, but sadly infeasible in most other states—and impossible to use in a districting algorithm.

Might a less manual process be feasible for automated districting algorithms? Nicholas Stephanopolous (2012) used “a wide array of demographic and socioeconomic information from the American Community Survey (ACS)” (p. 283) to find the “best available proxies for how closely the districts correspond to geographic communities of interest”. Through factor analysis, he identified the following factors that best predicted residential location throughout the nation: income, education and profession, marital status, race, age, and the agriculture-service divide. He then identifies communities of interest as contiguous groups of Census tracts that are very similar on a weighted average of the factors.

Stephanopolous’s work is excellent but falls short. Most prominently, because Stephanopoulos uses a weighted average of all the composite factors, he will not be able to pick up district heterogeneity. Certain factors will be more salient in some regions than others. Consider the agriculture-service divide. The overall weight Stephanopoulous assigns to this divide is very low, because it explains very little variance in most Census tracts: it is irrelevant in metro areas. Yet it is vital in exurban and rural areas. This means that his measure will fail to distinguish some communities of interest in exurban and rural areas.

This is not merely an econometric corner case. In fact, Stephanopolous himself states that the factors and weights can vary greatly: “Depending on whether I examined a nation as a whole or specific states, or entire districts or just their minority populations, the factor analysis condensed the . . . variables into anywhere from five to eight composite factors. . . The analyses of individual states provided greater accuracy as to the composite factors that matter most for those particular jurisdictions” (p. 1941). Using more fine-grained data can ameliorate this concern, but even within a state there can be great heterogeneity.

More fundamentally, moreover, there is a key difference between “universal” communities of interest and “identifying” ones. There are communities of interest that depend on your identity: your race, religion, or station in life. But there is a more agnostic—and universal—community of interest that is formed simply between people that live near one another. That is, regardless of their race, age, or socioeconomic background, neighbours use the same fire station, police force, and town hall; go to the same malls; walk down the same streets; and so on. While it might still be good overall to district for “identifying” communities of interest, there is a worry that districters might promote (and in so doing ossify) existing social cleavages, or unfairly privilege particular communities of interest. Using compactness is advantageous in that it measures universal communities of interest while remaining agnostic to identifying ones.

And finally—even if we could compute communities of interest well enough, and agree that it should be included in automated districting algorithms, we should nonetheless include a measure of compactness. This is because geographic compactness and communities of interest are separate concepts, and can come apart. One could draw a compact district that splits neighbourhoods and districts in twain, or a highly noncompact district that incorporates pockets of similar, yet isolated, communities throughout the country.

The courts are well aware of the distinction. In one of their judgments the Supreme Court mentions two black populations being “260 miles apart in distance and worlds apart in culture” — but if compactness were only important as a proxy for communities of interest, the Court would not have mentioned both distance and culture. The Californian Redistricting Commission’s final report also lists “geographic integrity” (communities of interest) and geographic compactness as separate criteria.

## If travel times is really better, why hasn’t prior work used it?

Previous work has talked about the benefits of using travel times, namely that it accounts for political geography and better captures communities of interest. Fryer and Holden Jr (2011) created a compactness index based on Euclidean distances, but wrote about its shortcomings (it was an imperfect proxy for actual travel distances). In fact, they specifically called out driving times and communities of interest as a natural extension of their work: “one can extend much of [our analysis] by using driving distance or what legal scholars refer to as ‘communities of interest’ ”.

In calculating the index, they used Euclidean distance, and wanted to "incorporate more general notions of distance into an empirically tractable algorithm". Sadly, they could not do so, as their compactness metric required finding the "maximal compactness" of a state. This is an non-polynomial (NP-hard) problem, which means that the computation needed to solve it is orders of magnitude greater than what we could possibly bring to bear. Similar concerns were echoed by Brian Olson, the creator of BDistricting, who also chose to use Euclidean distances rather than travel time because: "It might be the right kind of thing to measure, but it would take too long… the large amount of map data and extra computer time to calculate all those travel times would slow the process down horribly. It would then require a room filling supercomputer to get an answer in a reasonable amount of time".

Therefore, while many have recognised the theoretical advantages of using travel times over Euclidean distances / geographic compactness, no one has of yet come up with a computationally feasible way to use it. By adopting techniques like downsampling, memoisation, and the use of data structures from computer science, I can make the calculation of travel times computationally feasible, and (hopefully) competitive with existing compactness algorithms like Convex Hull or Polsby-Popper.

Additionally, in the course of my work I will build a database of travel times between voters, which can be used in empirical work as a drop-in replacement for Euclidean distances.

# Proposed metric

I propose a new metric, which I call *human compactness*.

Consider the sum of travel times from a voter $i$ to his k-nearest neighbours, where k is the number of voters in a voter's district. Call this sum $S_{i,knn}$, or the nearest-neighbour sum.

Consider the sum of travel times from a voter $i$ to all the other voters in his district. Call this sum $S_{i,kd}$, or the district sum.

Human compactness is defined as the ratio of the nearest-neighbour sum to the district sum for all voters. That is,

$$HC = \frac{\sum_i^K S_{i,knn}}{\sum_i^K S_{i,kd}}$$

A district is more humanly compact the more a voter's nearest neighbours correspond to a voter's actual co-districtors.

Like other commonly used metrics of geographic compactness, human compactness takes values from 0 to 1. Because the metric takes a ratio, this does not unfairly penalise states which are sparsely populated.

[diagram here...]

This metric works with a variety of different automated districting algorithms. It can be used with Monte Carlo Markov Chain approaches adopted by the Metric Geometry and Gerrymandering Group, as well as the graph partitioning approach by Levin and Friedler (2019). # TODO actually the graph partitioning approach doesn't take into account anything afaik

## Calculating the metric

### Pre-calculation step (things that can be done in advance)

1. Generate voter representative points (VRPs) by downsampling voter data
2. Calculate a table of driving times between all VRPs ($O(n^2)$)
3. For each VRP, calculate the nearest-neighbour sum: that is, sum the driving times of each VRP's k-nearest neighbours.

### Actual calculation step

Given a districting plan (a list of coordinates of polygonal shapes), do the following:

1. For each district. find the VRPs that fall inside it ($O(n \log n)$)
2. For each VRP in the district:
    1. Look up its nearest-neighbour sum ($O(1)$)
    2. Calculate its district sum by summing the driving times between it and all other VRPs in the district ($O(k)$)
    3. Add both sums to a running tally

3. Return the ratio of the sum of all nearest-neighbour sums to the sum of all district sums as a district human compactness score.
4. Return the total sum of all districts' human compactness scores.

## Complexity analysis

We don't really care about the pre-calculation step as that computation can be done in advance.

If we assume that travel times are symmetric, then the travel times from each VRP to all other VRPs in the districts are identical, and we'll only need to calculate the district sum once for each district. This gives a runtime dominated by the first step of $O(n \log n)$ time.

But if travel times are not symmetric (one-way roads), then each VRP has a different district sum, which naively takes $O(k)$ time per VRP, which means a total complexity of $O(k^2)$ time.

## Real-world analysis