

PV251 Visualization

Autumn 2024

Study material

Lecture 6: Geospatial data visualization

Geospatial data differs from other types of data mainly because it describes objects or phenomena that occur in the real world. Geospatial data appears in several applications, such as credit payment systems, telephone networks, censuses, etc. In this lecture, we will focus on an overview of special features and methods that are used to visualize geospatial data. This area is often referred to as geovisualization. We will show the most important basics of geospatial visualization, such as maps, and introduce visualization techniques for point, line, area, and surface data.

This area is widely covered by GIS (geographic information systems) and cartography, so here we will deal with this area purely from a visual point of view. After this lecture, we should have a general overview of state-of-the-art visualization techniques for geospatial data, and we should be able to implement and use them.

Large sets of spatial data are often created by the accumulation of discrete samples of a continuous phenomenon in the real world. Currently, there are several applications for which it is very important to analyze and display relationships between data that include geographic locations. Examples are modeling of global climate development (e.g., measuring temperature, precipitation, wind speed), monitoring economic and social indicators (unemployment rate, level of education, etc.), customer behavior analysis, telephone call statistics, credit card payments, or crime statistics.

Thanks to the special properties of such spatial data, their basic visualization is straightforward - spatial attributes are mapped directly to two dimensions of the output device (screen), thus achieving a map display.

Points, lines, areas

Maps can simply be considered a representation of the world, which is reduced to points, lines, and areas. Visualization parameters, such as size, shape, value, texture, color, orientation, supplement the displayed data with additional information.

According to U.S. Geological Survey, map visualizations are defined as a set of points, lines, and areas, which are defined by their position in the coordinate system (spatial properties) and by other "non-spatial" attributes.

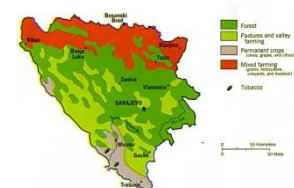
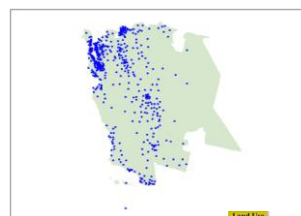
It is clear from the definition that we can distinguish spatial phenomena according to their spatial dimension into:

- **Point phenomena** - do not have a spatial component. They can be marked as 0-dimensional and can be specified using a pair (longitude, latitude) and a set of other attributes. Examples are buildings, oil wells, cities, ...
- **Line phenomena** - have a certain length, their width is given by default. They can be marked as 1-dimensional and can be specified using a set of pairs (longitude, latitude). Examples are large telecommunications networks, roads, borders between states... Attributes associated with bar phenomena may include capacity, traffic congestion, names,...
- **Area phenomena** - contain length and width. They can be marked as 2-dimensional and can be specified using a set of pairs (longitude, latitude) that are enclosed in a given region. Each pair can again have an associated set of other attributes associated with it. Examples are lakes, political map - states, ...
- **Surface phenomena** - in addition to length and width, they also contain height. Thus, they are referred to as 2.5-dimensional and may be specified by a set of vectors composed of longitude, latitude, and altitude, and each pair (longitude, latitude) may have additional attributes associated with it (altitude).

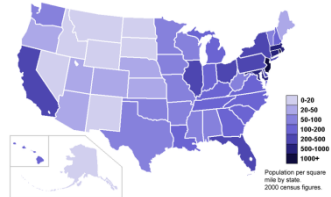
Types of maps

Maps can be divided according to their types based on the properties of the displayed data (qualitative vs. quantitative, discrete vs. continuous) and based on the properties of the so-called graphical variables (points, lines, surfaces, volumes). Examples of the resulting maps can be:

- Symbol maps (nominal point data)
- Point maps (ordinal point data)
- Land use maps (nominal area data)



- Choropleth maps - used to display phenomena using colored areas and shades. For example, population density is displayed.



- Line diagrams (nominal or ordinal line data)

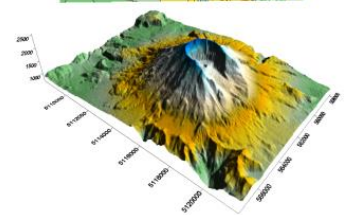
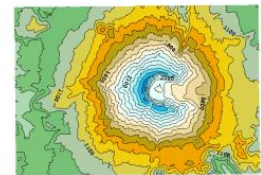


- Isoline maps (ordinal surface data)



- Surface maps (ordinal volumetric data).

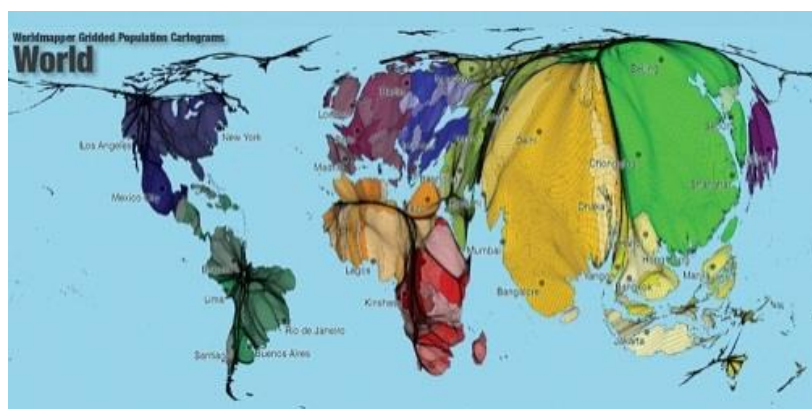
In addition, the image shows the same data displayed with a contour and then with a surface map.



Different types of representations

It is clear from the examples that the same data can be displayed using different types of maps. By aggregating point data within areas, we can create a choropleth map from a point map. Similarly, a land use map can be created from a symbol map. We can also generate a surface with a point map density display and display it as an isoline map or a surface map.

If we group point data inside areas and map the number of points inside a given area to their size, we get a so-called cartogram - thematic map. The picture shows the cartogram of the world population.



Exploratory geovisualization

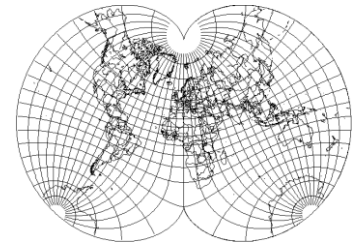
In exploratory geovisualization, the ability to interact with maps is crucial. Compared to traditional cartography, the classification and mapping of data using this technique is interactively adapted to the needs of the user. At the same time, interactive queries are enabled. These interaction capabilities are supported by a set of current techniques and systems. These allow, for example, the connection of several maps or a combination of maps with standard statistical visualization, such as bar and line graphs. Maps can also be combined with much more complex techniques for multidimensional visualizations, such as parallel coordinates.

Map projection

Map projection plays a key role in the visualization of geospatial data. Map projection deals with the mapping of positions on the globe (spheres) to positions on the screen (planar surface). Map projection is defined as $\Pi: (\lambda, \phi) \rightarrow (x, y)$. The data format for degrees of longitude (λ) is limited to the interval $[-180, 180]$, where negative values indicate western longitude and positive values represent eastern longitude. Degrees of latitude (ϕ) are defined similarly for the interval $[-90, 90]$, where negative values represent south latitude and positive values represent north latitude.

Map projections can have different properties:

- **Conformal projection** - correctly preserves the local angles of each point on the map. This means that it also preserves shapes locally. However, the area is not preserved.

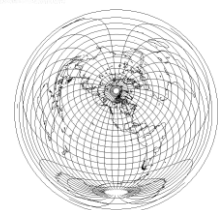


- **Equivalent projection** - a specific area of a part of the map covers exactly the same surface on a sphere. The resulting map distorts the shape and angles. For example, a square on the surface of a globe is mapped to a rectangle on a map of the same size.

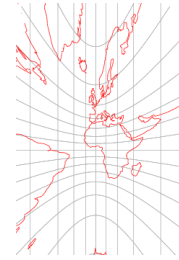


The Lambert planar equal-area projection is mathematically derived to display the property of equivalence.

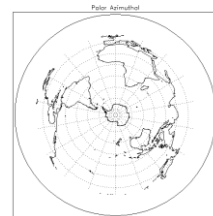
- **Equidistant projection** - maintains the distance from any point or line.



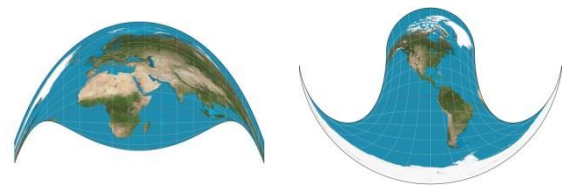
- **Gnomonic projection** - allows the display of "great circles" using lines. The main circle divides the sphere into two equally large hemispheres (at the globe it is the equator). Gnomonic projections maintain the shortest path between two points. The entire hemisphere cannot be displayed because the edges "run" indefinitely.



- **Azimuthal projections** - maintain the direction from the center point. Usually, this type of projection has radial symmetry, for example the distances from the center point are independent of the angle and at the same time concentric circles centered at the center point of the projection are projected on the circles centered at the center point of the map.



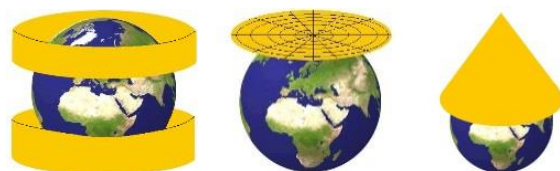
- **Retro Azimuthal projection** - the direction from point A to a fixed location L corresponds to the direction from S to L on the map.



Map projections - classification according to the type of surface

Map projections can also be classified according to the type of surface on which the sphere is projected. The most important types of such surfaces are:

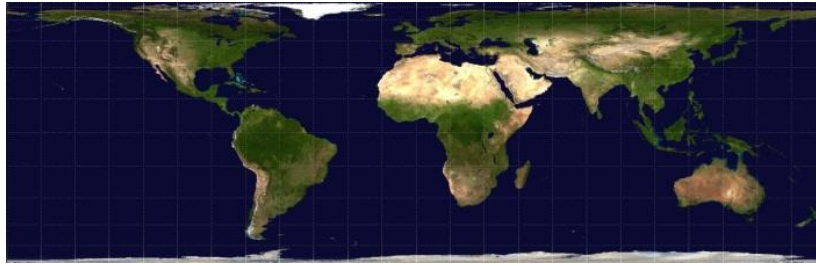
- Cylindrical projection
- Planar projection
- Conical projection



Cylindrical projection

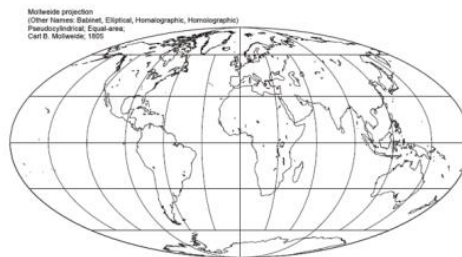
A cylindrical projection projects the surface of a sphere onto a cylinder placed around this sphere. Each point of the sphere is then projected onto the outer cylinder. Cylindrical projections have the advantage of being able to display the entire spherical surface. Most cylindrical projections maintain local angles and are therefore conformal.

Degrees of longitude and latitude are usually orthogonal to each other.



Pseudo-cylindrical projection

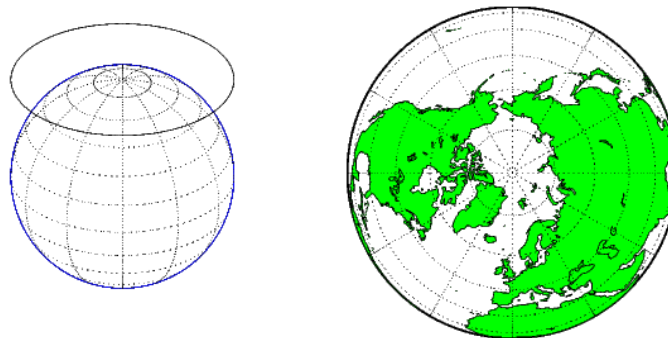
Pseudo-cylindrical projections represent the main meridian and each parallel as one straight line, the other meridians are deformed.



Planar projection

Planar projections are azimuthal projections mapping the surface of a sphere to a plane that is tangent to that sphere. The tangent point corresponds to the center of the projection.

Some planar projections are perspective.



Conical projection

The principle of conical projection is the mapping of the surface of a sphere to its tangent cone. Degrees of latitude are represented by circles centered on the projection center, degrees of longitude are straight lines emanating from that center.

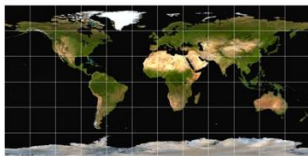
Conical projections can be designed to maintain distances from the center of the cone. There are also a number of pseudo-conical projections that preserve, for example, distances from the pole and at the same time distances from the meridians.



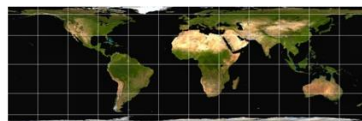
Examples of commonly used map projections

We will now list several widely used map projections. The variables defined in the map projections are listed in the following table.

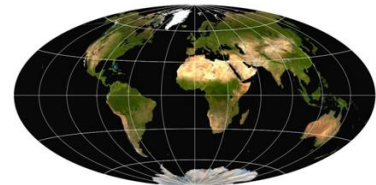
φ	measured degrees of latitude in radians
λ	measured degrees of longitude in radians
x	horizontal axis of the two-dimensional map
y	vertical axis of the two-dimensional map
$\varphi_0; \lambda_0$	latitude of the standard parallel resp. meridian measured in radians



Equirectangular



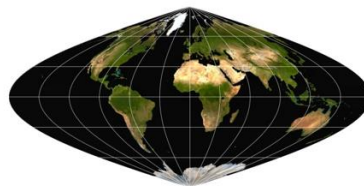
Lambert cylindrical



Hammer-Aitoff



Mollweide



Cosinusoidal



Albers equal-area conic

Visual variables for spatial data

Maps are used in different ways - for example, to find out information about a certain location, to find out general information about spatial patterns in many maps. Therefore, the mapping of spatial data properties to visual variables must reflect these goals.

The visual variables for spatial data are as follows:

- **Size** - size of individual symbols, width of lines, ...
- **Shape** - the shape of individual symbols or patterns in lines and areas

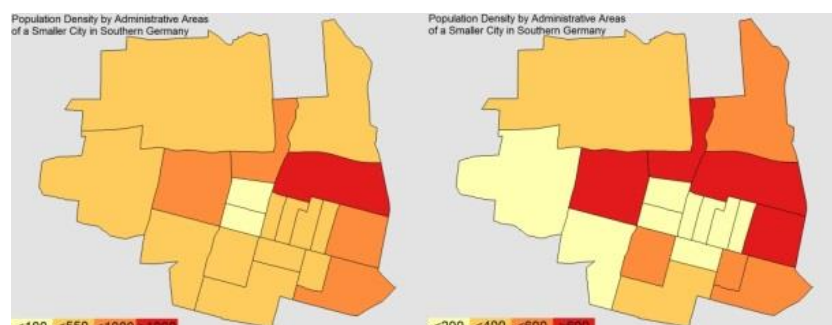
- **Brightness** - brightness of symbols, lines, and areas
- **Color** - color of symbols, lines, areas
- **Orientation** - orientation of individual symbols or patterns in lines and surfaces
- **Texture** - placement of patterns in symbols, lines, and areas
- **Perspective height** - perspective 3D view of phenomena where data values are mapped to the perspective height of individual points, lines, and areas
- **Arrangement** - arrangement of patterns in individual symbols (for point phenomena), patterns of dots and dashes for line phenomena and patterns of regular vs. random distribution of symbols for surface phenomena

	Size	Shape	Brightness	Color	Orientation	Spacing	Perspective height	Arrangement
Point								
Linear								
Areal								

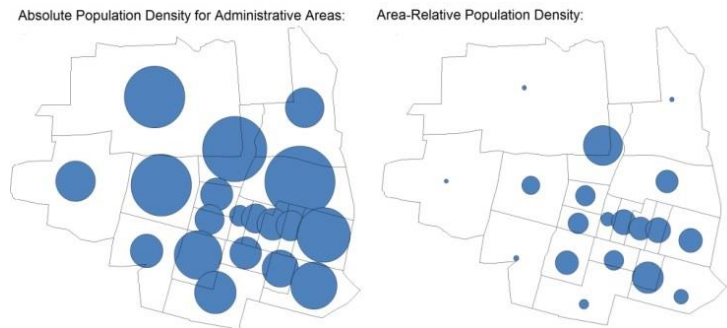
The effect of editing the input data on the resulting map

Cartographic designs have been intensively studied for several decades, during which time high-quality guides to these map designs have been created. All are based on the results of research in the field of human perception. In addition, the input data is often subject to various modifications (sampling, segmentation, normalization, ...), which can have a significant effect on the resulting visualization.

As an example, let's take two visualizations of the same data in the form of a choropleth map. The only difference is the different choice of departments of individual "classes", which has a very significant impact on the generated maps.



Another example is a significant change in the resulting visualization, which was created only by changing the absolute mapping to relative. Absolute numbers are visualized on the left, while they are displayed on the right relative to the size of the population. Due to large differences in the population in some areas, the expression may be quite the opposite of that in absolute numbers.



Visualization also strongly depends on the boundaries of the areas we cluster. The picture shows a well-known case of London cholera with different ways of clustering areas, leading to different choropleth maps.

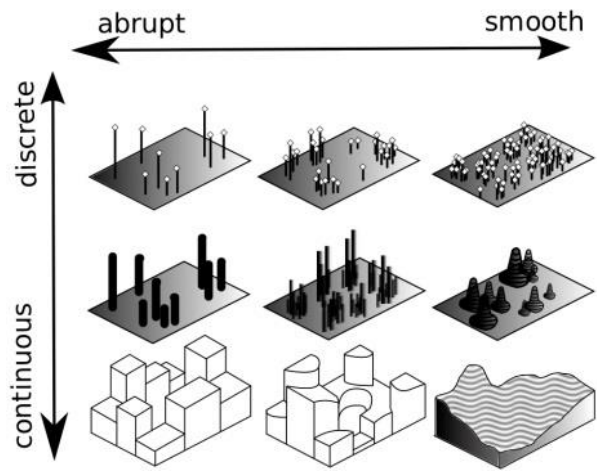


Geospatial data visualization

In the next part we will focus on visualization techniques for data according to their types. We focus on three basic types of data - point, line, and area.

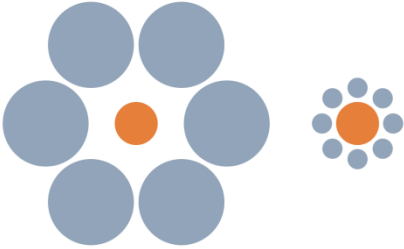
Visualization of point data

The first important class of spatial data is point data. They are discrete in nature, but may describe a continuous phenomenon, such as the measurement of temperature at a given location. Depending on the nature of the input data and the task required, the designer must decide whether he wants to display the data continuously or discretely, smoothly, or intermittently. The figure illustratively shows possible combinations of these decisions. For discrete data, we assume that they occur in certain places, while continuous data must be defined in all places. Smooth data changes gradually, while abrupt changes suddenly.



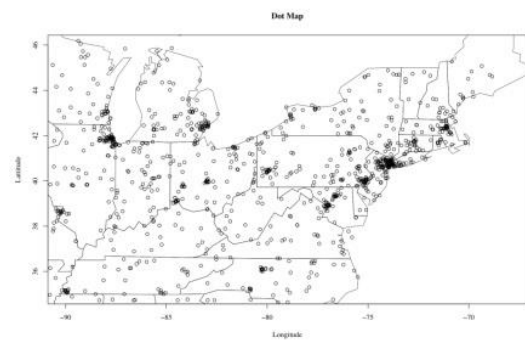
Point phenomena can be visualized in such a way that the given symbol is placed at the place where the phenomenon occurs. The simplest visualization of this type is called a point visualization. The quantitative parameter can be mapped to the size or color of the symbol. The most commonly used symbols in point maps are circles, but it is possible to use squares, columns, etc.

If the size of a symbol is associated with a quantitative parameter of a point, it is necessary to consider how to scale the symbols. Namely, the correct calculation of the size of individual symbols does not necessarily mean that the symbols will also be perceived correctly. The perceived size of the symbols does not necessarily correspond to the actual size, mainly due to problems with human perception - the perception of the size of the symbols depends on the local environment. Therefore, there is no global calculation for perceptual size perception.



The figure shows an example of different perceptions of the size of a circle depending on the local environment. The phenomenon shown is called the Ebbinghaus illusion.

Similarly, if color is used to represent a quantitative parameter, we must similarly take into account color perception problems.



Point maps are an elegant way to communicate a large amount of information about the relationships between point phenomena in a compact, convenient, and familiar form. However, when displaying large amounts of data on a map with different data densities, overlaps may occur in densely populated areas (e.g., populations), while sparsely populated areas remain essentially empty (see figure). The figure on the left represents the spatial distribution of the event. If we zoom in on the map, it can be seen that there is a considerable overlap of data (figure on the right side). Examples of this type of spatial data are credit card payments, telephone calls, health statistics, environmental records, crime statistics, etc.

It is worth noting that analyzes can contain several parameters that can be plotted on several maps. If all these maps present data in the same form, it is possible to put individual parameters into certain relationships and detect local correlations between data, dependencies, and other characteristics.

There are several approaches to solving the problem of displaying "dense" data. One of the widely used methods is 2.5D visualization clustering data points into regions. This technique is available in commercial systems, such as Visual Insight's In3D or ESRI's ArcView.

An alternative approach showing more detail displays individual data points as columns relative to their statistical value on the map. This technique is used by systems such as MineSet from Vero Insight and Swift 3D from AT&T. The problem with this method is that the columns overlap in the case of large datasets. In the end, therefore, only a certain part of the input data is visible.

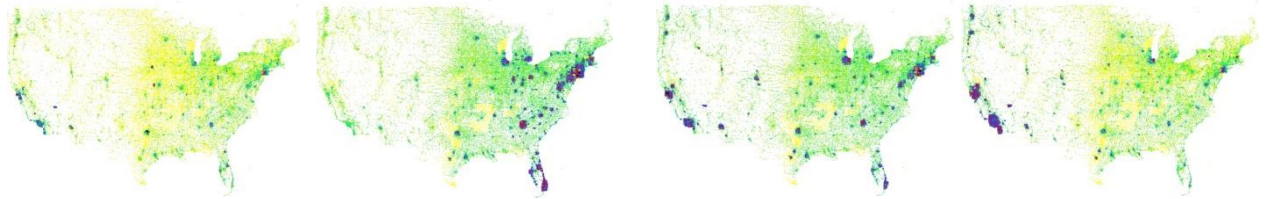
PixelMaps

An approach to avoid clustering data but also solve the overlaps is the PixelMap approach. The main idea is to move pixels that would otherwise overlap. The relocation algorithm recursively divides the dataset into four subsets containing data points in four equally large subregions. An efficient implementation of this algorithm uses a structure based on a quad-tree approach that supports a recursive division process.

The division process works as follows. We start at the root of the quad-tree and in each step, we divide the data space into 4 subregions. The condition for division is that the space contained in the subregion (in pixels) is greater than the number of pixels belonging to the

given subregion. If, after several repetitions of recursion, only a limited number of data points remain in the subregion, the points are placed using the so-called "pixel placement" algorithm, which places the first data item in its correct position and subsequent data items are placed in the nearest unoccupied positions. The resulting location is locally quasi-random.

The problem with displaying point data using PixelMaps is that in areas with high data overlap, the relocation of individual points depends on the order in which they are stored in the database.



The figure shows the four time slots displayed using this type of visualization (0:00 AM, 6:00 AM, 10:00 PM, and 6:00 PM, all EST time zone), when the volume of telephone calls in the United States was recorded at ten-minute intervals. The visualization intuitively shows the "development" of the volume of telephone calls according to time zones – i.e., when people wake up in the given areas or, for example, the decline of calls at lunch time. The visualization reveals both the expected patterns of behavior and the unexpected ones - for example, where the largest call centers operating overnight are located.

Line data visualization

The basic idea is to visualize spatial data describing linear phenomena to represent them using line segments between two endpoints determined by longitude and latitude. Standard line data mapping also allows you to map other input data parameters, such as line width, line pattern, color, lines, and icons.

In addition, it is possible to map start points, end points, and intersections to nodes with a specific color, size, shape, and label. The lines do not have to be straight, they can be polylines or splines.



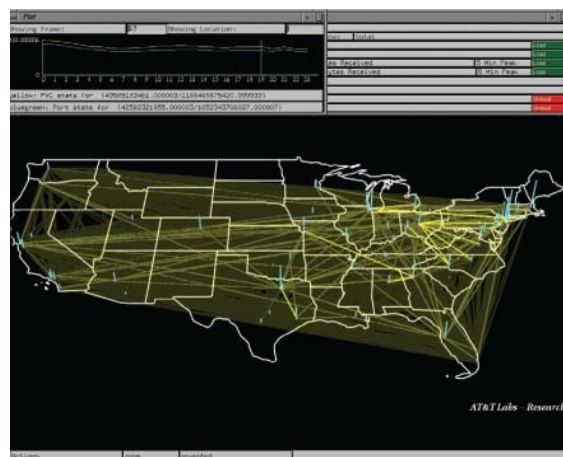
Network maps are used in a wide range of applications. Some approaches only show network connectivity to understand their structure. Eick and Wills used features such as aggregation, hierarchical information, node positions, and more to explore large networks with a hierarchical structure and no natural arrangement. They also used the color and shape to encode the information in the nodes and the width and color of the line to encode the connection information between the nodes.

The NCSA research group added 3D graphics to these network maps to animate the movement of packets in the Internet backbone network.

Becker, Eick and Wilks described a system known as SeeNet. The main idea is to involve the user in the process and let him interactively control the display and thus focus on interesting patterns in the data. To do this, they use two static views of networks to visualize geographic relationships.

Another interesting system for visualizing large network data is the Swift-3D System of AT&T (see figure), which integrates a set of relevant visualization techniques.

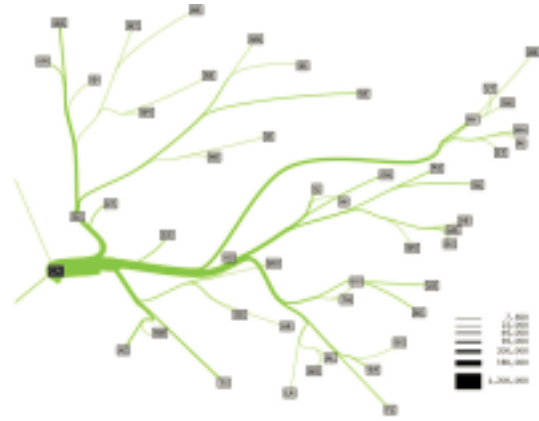
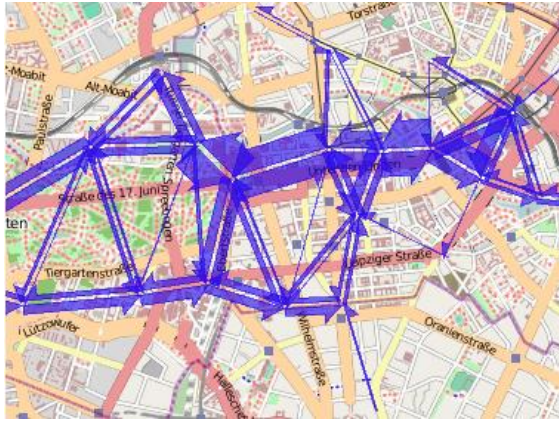
All of the network display techniques mentioned suffer from the problem of overlapping line segments in areas with dense data coverage.



Flow maps

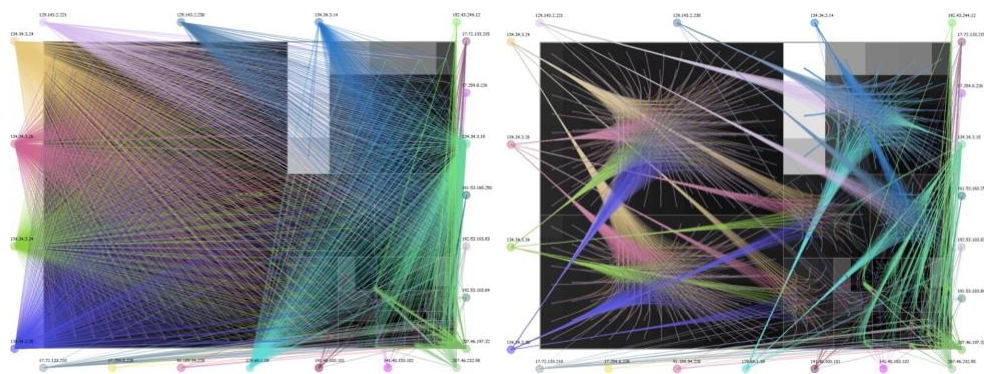
The flow map technique is inspired by graph algorithms that minimize edge intersection and deformation of node positions while maintaining their relative position.

The algorithm is based on hierarchical clustering according to the position of nodes and flows between them, which can compute clusters and redirect flows. Compared to other computer-generated flow maps (see image on the left showing the flow of tourists around Berlin), Stanford flow maps can produce much "cleaner" graphs (see image on the right showing migration from California).



Edge clustering also helps reduce ambiguity in line representation. If we have a hierarchy defined above the nodes, the corresponding edges can be clustered according to it. Nodes connected through the root of the hierarchy are maximally bent, while nodes at the same level of the hierarchy are bent minimally. This method can be combined with a number of other visualization techniques, such as traditional maps, tree visualizations, etc.

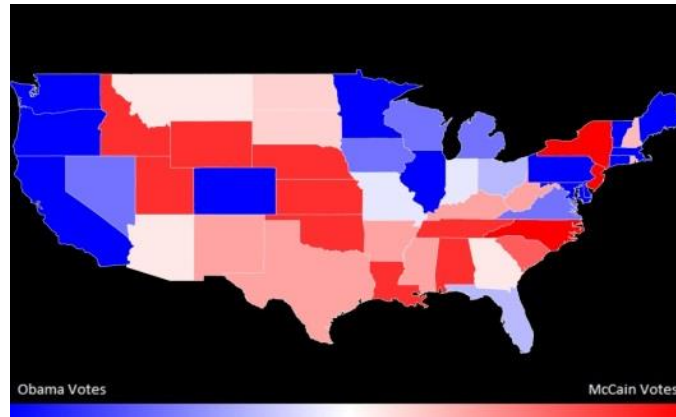
The figures show edge clustering applied to visualize IP flow traffic. The visualization shows traffic from external nodes to internal nodes displayed using a treemap. The picture shows the advantage of displaying edge clusters (right) compared to the original display of all straight edges (left).



Visualization of area data

The so-called **thematic maps** are most often used for the visualization of surface phenomena. These maps have several variants. The most popular variant are the so-called **choropleth maps** (Greek choro = area, pleth = value), where the values of attributes are coded using colored or shaded regions on the map.

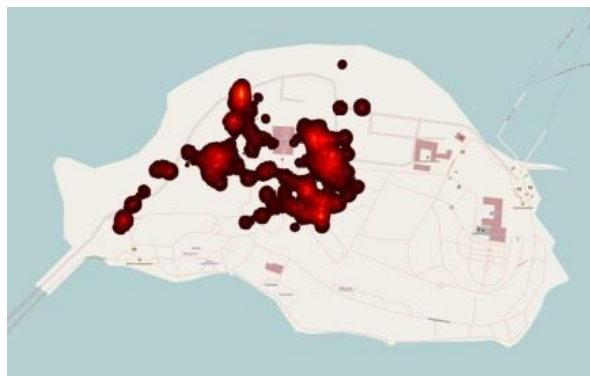
Choropleth maps assume that the mapped attributes are uniformly distributed within the regions. The picture shows a sample choropleth map showing the results of the 2008 election in the United States (Obama vs. McCain).



If the attribute has a different distribution than the division into regions, it is necessary to use other techniques, such as so-called **asymmetric maps**.

In asymmetric maps, the displayed variable forms an area independent of the original regions. For example, compared to choropleth maps, the area boundaries derived from an attribute may not match the regions in the map.

Another important type of maps are the so-called **isarithmic maps** showing the contours of continuous phenomena (see the picture showing the concentration of photographs taken on the island of Mainau). Typical representatives of this type of map are contour and topographic maps.



If the contours are derived from real data points (for example, the temperature measured at a given location), then these maps are called **isometric**.

If the data is measured for a given region (for example, a region) and the data point is considered the center of gravity, then these maps show so-called **isopleth maps**.

One of the main tasks in generating **isarithmic maps** is to interpolate data points to obtain smooth contours. This can be achieved, for example, by triangulation.

A complex and less frequent mapping technique is the so-called **cartograms**, where the size of regions is scaled in order to display statistical information. Cartograms have several variants - ranging from continuous cartograms that preserve the topology of polygonal

meshes to discontinuous cartograms that simply scale each polygon independently of rectangular or circular approximations of surfaces.

Note that area information can also be visualized using discrete points or symbols on the map — for example, using symbols that are proportional to the statistical parameters on the map or a point map.

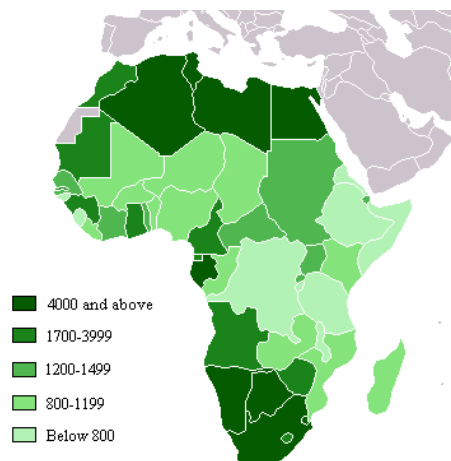
We will now focus in more detail on choropleth maps and cartograms.

Choropleth maps

Choropleth maps usually present surface phenomena in the form of shaded polygons, which are closed by a contour formed by a set of points, where the first and last points are identical. Examples are states, regions, parks, etc.

Choropleth maps are used to highlight the spatial distribution of one or more geographic attributes. When generating choropleth maps, data normalization (see the chapter on preprocessing of input data) and color and grayscale mapping are usually performed.

The problem with choropleth maps is that the most interesting values are often concentrated in densely populated areas (e.g., densely populated) with small and hardly visible polygons, and less interesting values are spread over sparsely populated areas, which are mostly represented by large and visually dominant polygons. Choropleth maps therefore highlight areas represented by large polygons, which are usually of lower importance.



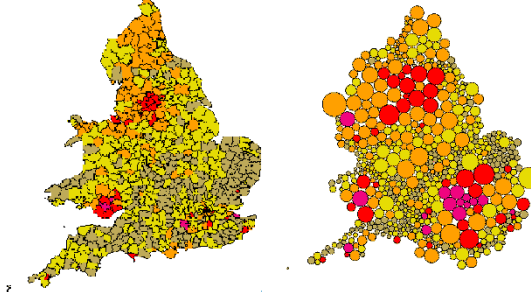
Cartograms

Cartograms are generalizations of common thematic maps, which try to avoid the problems that appear in choropleth maps, when geographical data is distorted based on their statistical value.

Cartograms are a specific type of map transformation, where the size of regions is changed based on a certain input variable, which is tied to the geographical properties of the input

data. Examples of the use of cartograms are the display of demographics, election results or epidemiological data.

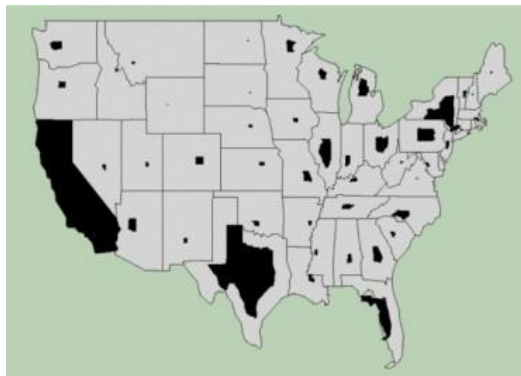
COMPARISON OF A TRADITIONAL MAP AND CARTOGRAM REPRESENTATIONS OF THE PERCENTAGE OF THE MALE POPULATION OF WORKING AGE IN 1891



Cartograms can be divided into several categories.

Noncontinuous cartograms

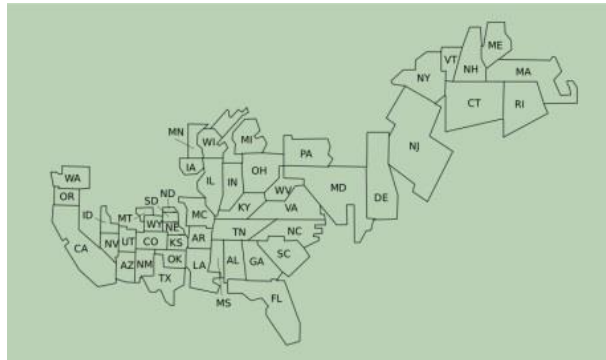
They fully comply with area and shape constraints, but do not maintain the topology of the input map. Because the scaled polygons are drawn inside the original polygons, there are no problems with the perception of such a map, despite the loss of topology. What's worse is that the original size of individual polygons limits their final size - so it is not possible to arbitrarily enlarge small polygons without scaling (enlarging) the entire map. As a result, important areas are difficult to see and screen space is very limited.



Noncontiguous cartograms

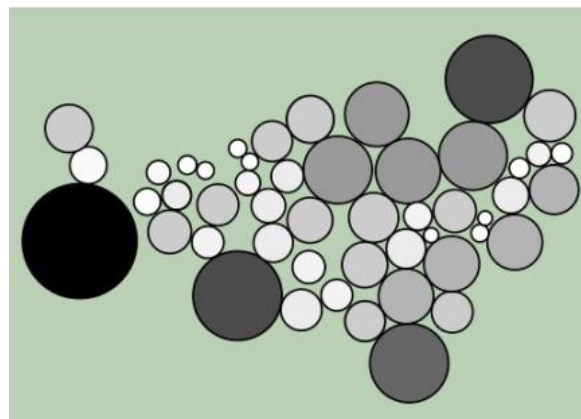
They scale all polygons to their target sizes, which exactly meet the space requirements. Individual shapes can be slightly relaxed, so the polygons touch, but without overlap. Thanks to this, the whole topology of the map is also highly relaxed, because the polygons do not maintain mutual "neighboring" relationships. This type of cartograms provides a very good arrangement of surfaces, including the preservation of their shapes. However, the global

shape and topology of the map is lost, which can worsen the perception of the map as a whole.



Circular cartograms

They completely ignore the shape of the input polygons and represent them with circles. In most cases, area and topological restrictions are also relaxed, so this type of cartograms shows the same problems as previous non-adjacent cartograms.

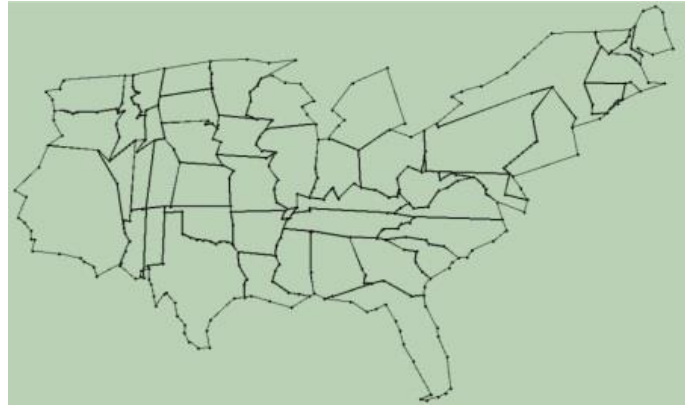


Continuous cartograms

The last category is the so-called continuous cartograms, which, unlike the previous types, completely preserve the topology of the map and

they relax given surface and shape restrictions. In general, cartograms cannot fully preserve shape and area, so generating cartograms involves a relatively complex optimization problem that seeks to find a satisfactory compromise between preserving shape and area.

Although it is difficult to generate continuous cartograms, the resulting polygonal meshes resemble the original map much more than other computer-generated variants of cartograms. Therefore, in the rest of this section, we will deal with continuous cartograms.

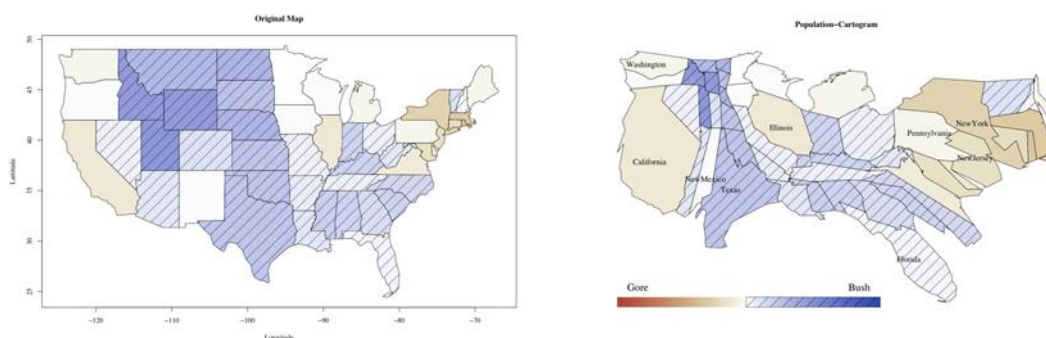


Generating cartograms

Creating cartograms manually is very difficult, so it is a very popular area to study automatic methods for generating cartograms using a computer. Cartograms can also be considered as a general technique for visualizing information. They provide a means to deal with shape retention versus surface retention by scaling the original polygons based on external parameters.

In the so-called "population" cartograms, more space is allocated for densely populated areas and these areas are additionally highlighted (because they most likely contain the most interesting data).

The image on the left shows a traditional map of the election results in the United States in 2000, the image on the right presents the same information using a population cartogram. In the cartogram, the area of the states is adjusted according to their population, which will allow to display the close election results in the individual states much more accurately and effectively than in the original choropleth map on the left.



To make a cartogram effective, its message must be quickly understood by the user, and at the same time the user must understand its relationship to the original map. This recognition depends in particular on the preservation of basic properties such as shape, orientation or adjacency. However, this maintenance is very difficult to achieve, especially for cartograms, and it has been shown that in general this problem of cartograms cannot be solved. Even if we allow the existence of these errors in the representation of shape and area, we still have

the problem of difficult optimization, thanks to which the current algorithms solving this problem are very time consuming.

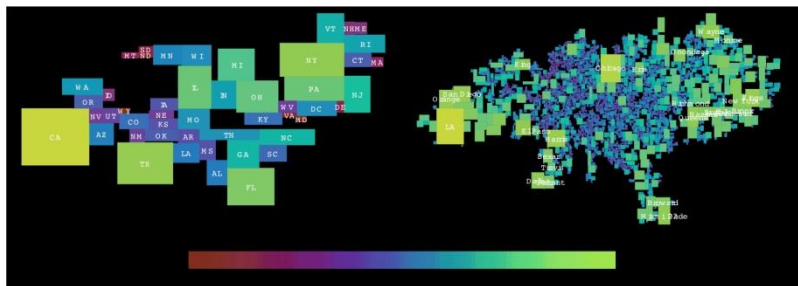
Rectangular cartograms

The main idea of rectangular cartograms is to approximate known maps covering a given area using rectangles and to find the distribution of available screen space, where the areas occupied by individual rectangles are determined proportionally to the statistical values.

To promote the best possible understanding of the information presented by such a cartogram, the rectangles are placed as close as possible to their original positions and original neighbors.

A problem can be defined as an optimization problem with a set of constraints and optimization criteria, including area, topology, relative position of polygons, rectangle size, and free space.

There are different algorithms for solving this problem. One of them is the so-called RecMap algorithm.



Map labeling

These are placed text or image tags near points, lines, and polygons. Although it seems like a simple and straightforward task, it has been proven not to be the case. There are a number of different algorithms that solve this problem, which differ in efficiency, quality of results. These algorithms are mostly based on heuristic methods.