Geospatial analysis in R: Part 2

Michael Otterstatter
BCCDC Biostats Session

Sept 13, 2019

Session overview

- Background and concepts
  - principles of data visualisation
  - review of map types
- Creating basic maps in R
  - `tmap`

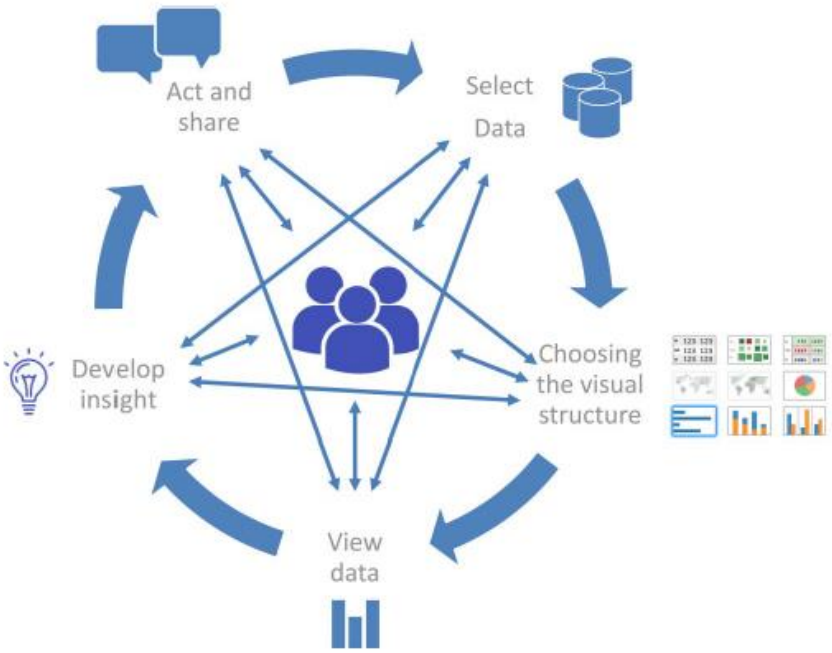
BACKGROUND AND CONCEPTS

Principals of data visualization

Data visualisation is

- A well-known component of scientific communication, but is increasingly recognized as a process unto itself for generating insights; and,

- a rapidly growing field, particularly with availability of large complex datasets and powerful viz tools

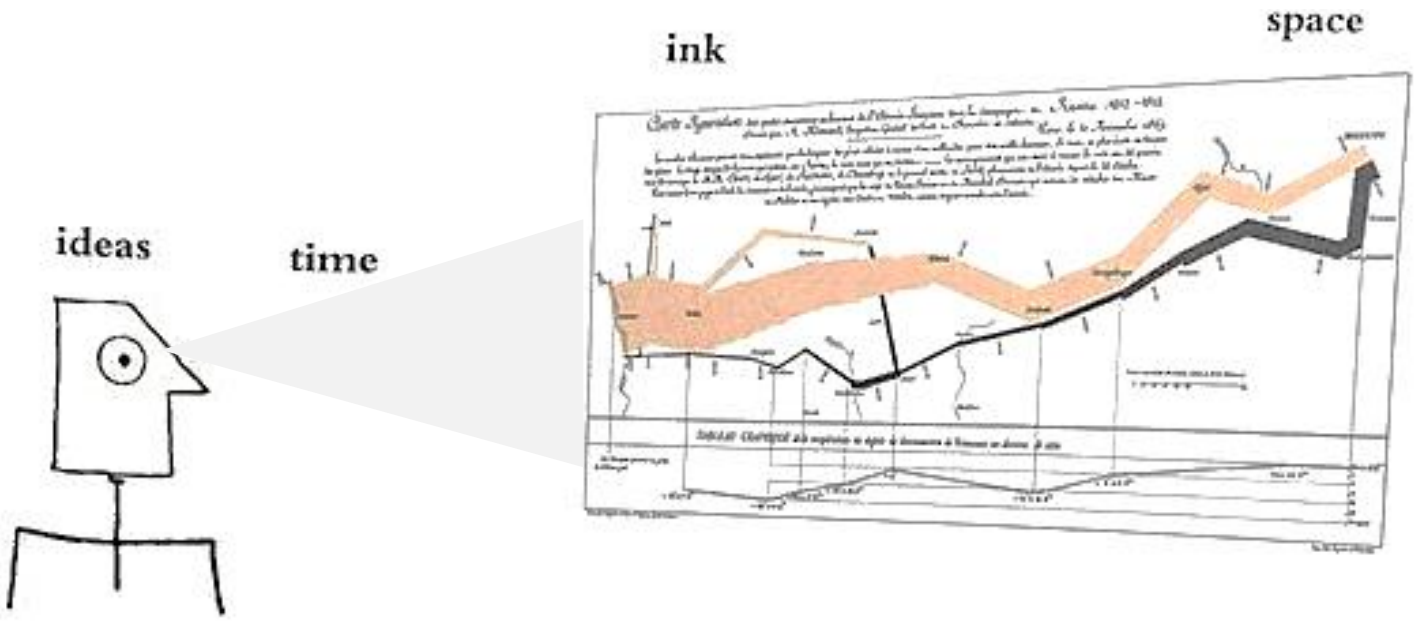


'The visual analytics cycle', Martinez et al. 2016

Principals of data visualization

- Consider a guiding principle in visualization…

  *"Graphical excellence gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space…and requires telling the truth about the data"**



*Tufte, 2001.

Principals of data visualization

…and a distinction, particular to maps, underscoring the need for accurate and effective map making:

> *"…maps differ from statistical graphics because the geographic setting portrayed in the map almost always triggers memories, opinions, and conclusions wholly separate from (but perhaps related to) the intent of the mapmaker"**

*Waller & Gotway, 2004.

Review of map types

- Point features vs aggregate features
  - *Point maps* represent specific geographic locations as dots (points) on the map, e.g., residence or facility, and are primarily used to show spatial distribution
  - A famous public health example is John Snow's maps of cholera cases (represented as points) – today, we often still show spatial distributions using simple dots on a map

Review of map types

- Point features vs aggregate features
  - Point maps can reflect a limited amount of attribute information through the use of differing symbols, and/or colours, sizes and transparencies of the points
  - Hence, point maps are most suitable when there is no attribute information shown, or when the attributes take on a small number of discrete values

Review of map types

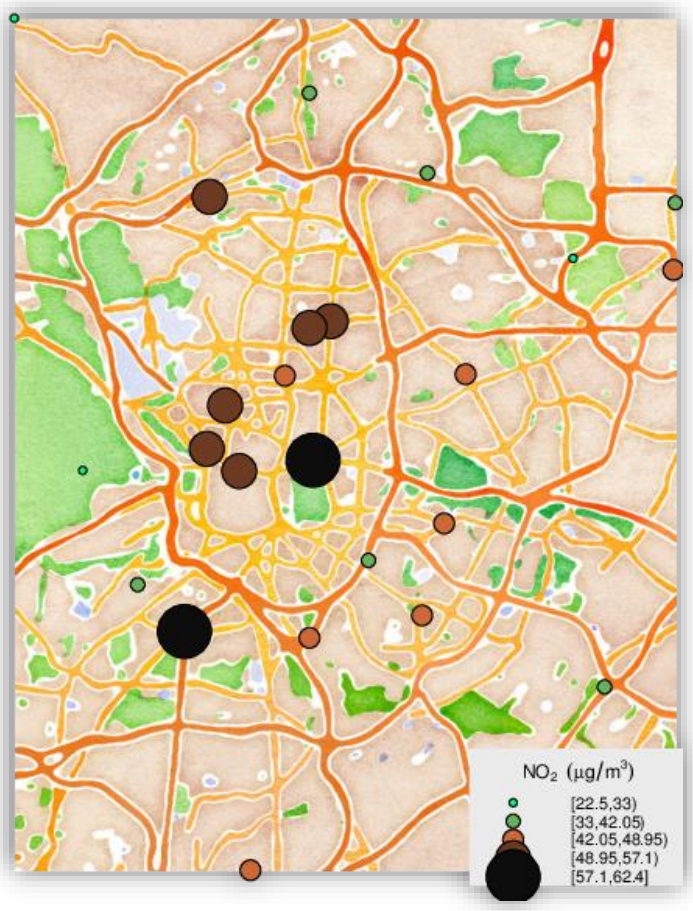
- Point features vs aggregate features
  - *Contour maps* are suitable when attribute information is a continuous range of values, e.g., topographic maps with isolines/shading to represent differing elevation

Review of map types

- Point features vs aggregate features
  - *Image maps* are based on pixels and variations in colour to show features, most notably aerial photographs and satellite imagery

Review of map types

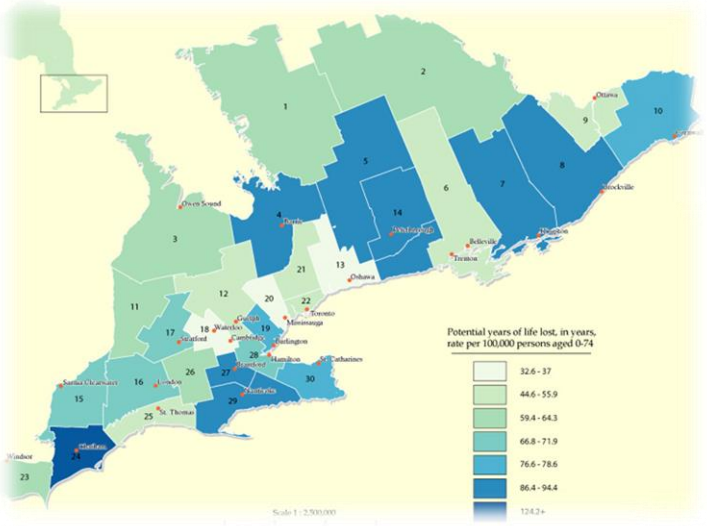
- Point features vs aggregate features
  - For areal data (areas and aggregations), maps use symbols, patterns or colour to represent attribute values for a region
  - For example, *proportional symbol maps,* use symbol size scaled proportionately to the value of an attribute for a given region



Lamigueiro, Ó. P. 2014

Review of map types

- Point features vs aggregate features
  - *Choropleth maps* display areal data by assigning colours to different attribute values and shading the regions of the map accordingly
  - Despite certain limitations (e.g., single colour implies constant regional rates), choropleth maps are common in public health given data are often available only as regional aggregates



Review of map types

- Point features vs aggregate features
  - Range of values shown in choropleth maps may be defined into classes (e.g., differing colours or patterns) of non-overlapping intervals in an attribute (‘classed choropleth maps’) or show a continuous change (e.g., differing hue or intensity of a single colour)

Review of map types

- So, which type of map to use? Consider:
  - Any working limitation or restrictions (e.g., only black and white)
  - appropriateness given the data at hand (point locations vs areal data)
  - In general, try a few different map types and see which best communicates the message in the data

Review of map types

- Reference maps vs thematic maps
  - *Reference maps* focus on locations and names of particular features, e.g., boundaries and names of standard geographic areas, and their major physical features, such as roads, coastlines, and bodies of water
  - *Thematic maps* focus on spatial variation in one or a few specific topics or 'themes' -- common types of thematic maps show distributions of population density or average income

* *Also see:* www150.statcan.gc.ca/n1/pub/92-195-x/92-195-x2011001-eng.htm4

Review of map types

- Reference maps vs thematic maps
  - Reference maps and thematic maps are not mutually exclusive: thematic maps typically contain some reference information, e.g., place names or roads, to help orient readers to the regions shown on the map

Review of map types

- Reference maps vs thematic maps
  - *Thematic maps* contain two key elements:

    i) a base map and ii) statistical data

    - Both elements usually come from digital files, e.g., cartographic boundary files and census data
    - often thematic maps are point maps or choropleth maps
    - thematic maps have a variety of purposes, including presentation of geographic data and exploratory spatial data analysis

Considerations and cautions

- Aesthetics (i.e., those aspects or principles concerned with perceptual value or beauty) are not covered here, but can be essential for accurate and effective communication with maps

CREATING BASIC MAPS IN R

Background

- Maps are one of the most fundamental and effective communication tools, and form the basis of geospatial analysis

- Map making is now more accessible through powerful and user-friendly software, such as R

- Although it is now easy to jump in and make maps, give careful attention to the accuracy and effectiveness of your maps as communication tools

Background

- Here, we focus on the R package `tmap` for modern, flexible mapping. For information on `tmap` see
  - https://cran.r-project.org/web/packages/tmap/ and
  - https://cran.r-project.org/web/packages/tmap/vignettes/tmap-getstarted.html

- The applications of `tmap` herein are based on Lovelace et al. https://geocompr.robinlovelace.net/adv-map.html

Background

- Static maps
  - Base R function `plot()` accepts vector and raster data and produces static maps quickly and simply; however, it is often not flexible enough
  - R package `tmap()` is a set of tools for creating thematic maps from vector or raster data
    - maps can be static or interactive maps
    - Syntax is concise, easy to learn, and works well with essential data manipulation tools in `tidyverse`

`tmap`

- like the primary plotting package in R (`ggplot2`), `tmap` is based on the principles of layer grammar of graphics (Wilkinson, 2012)

- Essentially `tmap` generates maps from a spatial dataset by forming one or more layers that each associate specific visual characteristics of map elements (‘aesthetics’, such as position, colour, size, transparency, etc.) with data variables

tmap

- To create a map, one uses the `tm_shape()` function plus (+)
  - one or more aesthetic layers, e.g., `tm_polygon`, `tm_fill`, `tm_symbol`, etc.

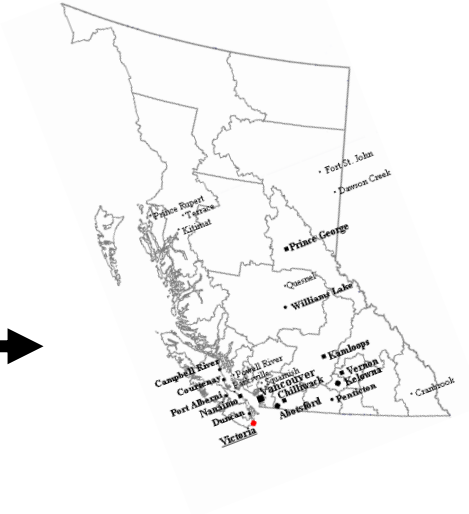
```
tm_shape(my_sf_data) +
   tm_polygon(col = "my_regions")
```

- `tm_shape()` specifies the sf spatial data object, whereas the layers specify how data variables are assigned to visual map elements (aesthetics).

dataset_a

Layer A [aesthetics: borders] →

Layer B [aesthetics: point locations] →




```
tm_shape(dataset_a) + tm_polygons() + tm_symbols()
```

dataset_b

→ Layer C [aesthetics: geography] →

```
tm_shape(dataset_b) + tm_fill()
```

`tmap`

- To see a list of the aesthetic layers that are available in `tmap`, type `help("tmap-element")`
- In order to generate a quick and simply thematic map, one can also use `qtm(my_sf_data)`

tmap

- In R, we typically assign data, variables, plots and all other things as 'objects'. Similarly, working with and saving maps is made easier by creating them as objects

```
my_base_map <- tm_shape(my_sf_data) +
      tm_polygon()
```

(e.g., starting with borders)

tmap

- New layers can then be added successively:

```
new_map1 <- my_base_map +
      tm_shape(more_sf_data) +
      tm_raster()          (e.g., adding elevation)

new_map2 <- new_map1 +
      tm_shape(even_more_sf_data) +
      tm_lines()           (e.g., adding bodies of water)
```

tmap

- Note that `tmap()` automatically uses certain default aesthetics (e.g., black lines, grey shading for fills, etc.), but that these are all modifiable in their associated layer.
- There are two types of aesthetics: i) static aesthetics – set visual aspect of map element to a constant value; ii) variable aesthetics – set visual aspects to vary according to values in the dataset

```
tm_shape(my_sf_data) +
   tm_fill(col = "red")

tm_shape(my_sf_data) +
   tm_fill(col = "my_landarea_variable")
```

`tmap`

- Aesthetics include elements such as colour (`col =`), transparency (`alpha =`), line width (`lwd =`) , line type (`lty =`), title (`title =`), etc.