UNIT 16 GEOGRAPHICAL DATA AND DATA MAPPING

Structure

16.0 Introduction

16.1 Learning Outcomes

16.2 Understanding Geographical Data

- 16.2.1 What is Geospatial Data?
- 16.2.2 Types of Geographic Data
- 16.2.3 Common Geographic Data Formats
- 16.2.4 Sources of Geographical Data for Journalists

16.3 The Power of Maps in Data Storytelling

- 16.3.1 Why Map Data? The Four Key Functions
- 16.3.2 Foundational Principles of Cartography

16.4 Common Types of Data Maps in Journalism

- 16.4.1 Choropleth Maps
- 16.4.2 Proportional and Graduated Symbol Maps
- 16.4.3 Point (or Locator) Maps
- 16.4.4 Heat Maps (Density Maps)
- 16.4.5 Flow Maps

16.5 A Survey of Popular Data Mapping Tools

- 16.5.1 No-Code Tools for Beginners
- 16.5.2 Low-Code Tools for Intermediate Users
- 16.5.3 GIS Software and Code Libraries for Advanced Users

16.6 Critical Considerations in Mapping

- 16.6.1 The Modifiable Areal Unit Problem (MAUP)
- 16.6.2 The Ecological Fallacy
- 16.6.3 How Classification and Colour Choices Create Bias
- 16.6.4 Privacy and Anonymity in Point Maps

16.7 Case Study: Mapping Election Turnout in India

16.8 Let Us Sum Up

16.9 Keywords

16.10 Further Readings

16.11 Check Your Progress: Possible Answers

ignou THE PEOPLE'S UNIVERSITY

16.0 INTRODUCTION

One of the most important questions a journalist can ask is, "Where?" The answer isn't just about a dot on a map. It is about providing context and finding hidden threads that connect people and events. When you are working with data, the map is your best tool for answering that question. These aren't just static pictures; modern data maps are living, breathing tools for investigation that can form the centrepiece of an investigation.

Think about it: you can track a disease spreading from one district to the next, show the gaps in infrastructure between states, or pinpoint air pollution hotspots in real time. Data mapping takes numbers and turns them into stories that people can see and relate to. It helps your readers understand how a big national story actually affects them in their own neighbourhood, making your work more personal and much more powerful.

This unit provides a comprehensive introduction to the world of geographical data and data mapping for journalists. We will start by breaking down what geospatial data actually is, where to find it, and what to do with it. Then, we will jump into the most common types of maps you will see in newsrooms, from the classic choropleth to more specialised visuals like flow and heat maps.

Crucially, this unit will equip you with a step-by-step workflow for creating a data map, from finding the right data to joining it with a map file and making critical design decisions. We will also navigate the significant ethical considerations inherent in mapping.

16.1 LEARNING OUTCOMES

After completing this unit, you will be able to:

- explain what geospatial data is and tell the difference between its types and formats,
- find and evaluate the quality of geographic and statistical data for a mapping project,
- understand why maps are so important for telling stories with data,
- know the difference between map types like choropleths, symbol maps, and point maps, and when to use each one,
- follow a clear workflow to make a basic data map by joining a spreadsheet to a map file, and
- pick the right mapping tool based on different levels of complexity and technical skill.

16.2 UNDERSTANDING GEOGRAPHICAL DATA

Before you can make a map, you need to get to grips with the special kind of data that makes it work. This is the world of geographical data, also called geospatial data.

16.2.1 What is Geospatial Data?

Simply put, geospatial data is any piece of information that can be tied to a specific place on Earth. It describes things, events, or features by their location.

Sometimes that location is explicit, like a latitude and longitude coordinate (e.g., 28.7041° N, 77.1025° E for Delhi). Other times, it can be implicit, like an address ("10 Downing Street, London") or just a name ("Mumbai"). For a computer to turn this into a map, that location info needs to be in a structured format it can read.

16.2.2 Types of Geographic Data

Geospatial data generally comes in two main varieties:

- **Vector Data:** This is the most common type used in data journalism. Vector data represents the world using basic geometric shapes: points, lines, and polygons.
    - **Points:** A single spot, marked by a latitude and longitude. Perfect for things like cities, hospitals, the centre of an earthquake, or a specific address.
    - **Lines:** A series of connected points. Used to show things like rivers, roads, railway tracks, or an aeroplane's flight path.
    - **Polygons:** A closed shape made of connected points. These represent areas with boundaries, like countries, states, districts, or parks.
- **Raster Data:** Raster Data uses a grid of tiny squares called cells (or pixels), and each pixel has a specific colour or value. This grid-based approach makes raster data perfect for showing things that don't have clear boundaries but instead blend into each other, like temperature, air quality, or elevation across a landscape.
    - **Examples:** Common examples are satellite photos, aerial imagery, and maps that show continuous information like temperature or rainfall.
    - While less common for direct use in typical news maps, raster data is often the source from which vector data (like the boundary of a flood zone) is derived.

16.2.3 Common Geographic Data Formats

When you go looking for map files, you will encounter a few common types:

- **Shapefile (.shp):** A very common vector format from a company called Esri. It is actually a collection of several files (.shp, .shx, .dbf, .prj, etc.) which have to be kept in the same folder since they contain information about different geographic features which is then put together and interpreted. It is a standard format supported by almost all Geographic Information System (GIS) software.
- **GeoJSON(.json or .geojson):** A modern, open-standard format for vector data. It is web-friendly, human-readable, and the preferred format for most interactive web mapping libraries. All the data is stored in a single text file.
- **KML (Keyhole Markup Language, .kml):** This was originally made for Google Earth. It is useful for sharing geographic data and can be dropped right into tools like Google My Maps.

- **TopoJSON:** This is an extension of GeoJSON but is much smaller in file size. It works by defining shared borders only once instead of repeating them, which makes it efficient for fast-loading web maps.

16.2.4 Sources of Geographical Data for Journalists

There are two kinds of geographical data that can be sourced by a journalist: geographic data which gives spatial information about geographical features (e.g. map shapes or road networks), and statistical data which gives numerical information or the numbers you want to visualise (e.g. population counts). These are most commonly found in the following sources:

- **Official Government Portals:** In India, places like the Survey of India and ISRO's Bhuvan portal are good sources for official administrative boundaries. Many government departments and data portals (like data.gov.in) also publish data that has location information.
- **OpenStreetMap (OSM):** This is like the Wikipedia of maps, a collaborative, open source map of the entire world. It is an excellent source to find detailed information on roads, buildings, and points of interest. You can pull data from it using tools like the Overpass Turbo API.
- **Natural Earth:** This is a public domain map dataset that provides free, high-quality global vector and raster data at various scales, perfect for country-level or world maps.
- **Humanitarian Data Exchange (HDX):** A great resource for finding geographic and statistical data related to humanitarian crises, development, and disasters.
- **Academic Institutions and NGOs:** These organisations often publish their research along with the datasets they use, which can include valuable geographic info.

Check Your Progress - 1

Note: 1) Use the space provided below for your answers.

2) Compare your answers with those given at the end of the unit.

1. What is the main difference between vector and raster data? Give an example of a feature best shown with each type.

   .....

   .....

2. You downloaded a map of Indian states that came in a zip file with files named states.shp, states.shx, states.dbf, and states.prj. What do you call this group of files, and why is it so important to keep them all together?

   .....

   .....

3. In mapping, what is the difference between geographic data and statistical data?

   .....

   .....

16.3 THE POWER OF MAPS IN DATA STORYTELLING

So far, we have explored what geographical data means. Now, we shall turn our attention to maps. Maps are not just navigational aids. They are powerful tools for analysis that can tell significant stories. Knowing why you are making a map is just as important as knowing how to make it.

16.3.1 Why Map Data? Four Key Functions

In journalism, a map usually does one or more of these four things:

1. **Show Location and Give Context:** The most basic job of a map is to show where something happened. A simple locator map for a news story gives the reader immediate context.
2. **Reveal Patterns and Distribution:** Maps excel at showing how things are spread out across an area. Is poverty clustered in one particular region? Are polling booths spread fairly across a constituency? A map can tell you all this.
3. **Compare Different Places:** A map is a very intuitive way to compare the same variable in multiple locations. For example, a map of literacy rates by state instantly shows you which state is ahead and which is falling behind.
4. **Show Relationships and Connections:** You can use maps to illustrate how places are connected, whether you are showing migration routes, trade flows, or the path a storm is taking.

16.3.2 Foundational Principles of Cartography

Put simply, cartography is the art and science of making maps. You don't need to be a professional cartographer to make a map, but you do need to know a few basic principles that will make your maps accurate and honest.

- **Projection:** A map projection is how you take the curved surface of the Earth and flatten it onto a screen or on a piece of paper. All projections distort reality in some way, whether in shape, area, distance, or direction. When you are making a map of the whole world, your choice of projection can make some continents look way bigger than they really are. This is less of a problem for local maps, but it is good to remember that no flat map is ever a perfect picture of the globe.
- **Scale:** Scale is the ratio between a distance on your map and the actual distance on the ground. A map of a single city is called "large-scale" because it shows a lot of detail. A map of the world is "small-scale" because it shows less detail. The scale you choose determines how much detail you can show.
- **Symbolisation and Generalisation:** A map is always a simplified version of reality. Cartographers use symbols (like points, lines, and colours) to stand in for real-world things. They also have to generalise, which means simplifying complex features to keep the map from getting cluttered. For example, a river with thousands of tiny bends might be shown as a single smooth line. The choices you make here have a huge impact on the story your map tells.

16.4 COMMON TYPES OF DATA MAPS IN JOURNALISM

Maps have a specific role in data storytelling, and so, picking the right kind of map is key to telling your story clearly and accurately. Here are the main types of maps you can use.

16.4.1 Choropleth Maps

This is the most widely used type of data map. A choropleth (from Greek words meaning "area" and "value") is a map where areas like states or districts are coloured in to represent the value of something. This is used to show rates or ratios, such as population density (people per square km), literacy rate, or per capita income. However, you must never use a choropleth map to show raw numbers (like the total population or the total number of crimes). Why? Because bigger areas will almost always have bigger numbers, so they will look darker on the map. Your map will end up being a map of area size, which will be misleading. You always have to normalise your data first by turning it into a rate or percentage (e.g., by dividing by the population or the area).

16.4.2 Proportional and Graduated Symbol Maps

These maps use symbols, usually circles, on top of different areas. The size of the symbol will change based on the value of your data. This is used to show raw counts or totals. For example, a map showing the total number of COVID-19 cases in major cities could use circles of different sizes for each city. This avoids the problem that choropleths have. In a true proportional symbol map, the area of the circle is directly proportional to the value. In a graduated symbol map, the data is grouped into classes (like 1-100, 101-500), and everything in the same class gets the same size symbol.

16.4.3 Point (or Locator) Maps

This is the simplest kind of map. It just uses dots to show the location of individual things. This is used to show the exact location of specific events (like the sites of recent protests), buildings (like all the public hospitals in a city), or just to give context for a news story (a "you are here" kind of map).

16.4.4 Heat Maps (Density Maps)

A heat map shows you the geographic density of your points. It uses colour to show where your points are most clustered together. This is used when you have so many points that plotting them all individually would just be a mess. A heat map can reveal the "hotspots" in your data. For example, you could map thousands of traffic accidents to see which intersections are the most dangerous.

16.4.5 Flow Maps

These maps use lines of different thicknesses or colours to show how people, things, or information move between places. This is used to show things like migration patterns, trade routes, air traffic, or how an idea spread on social media.

Check Your Progress - 2

Note: 1) Use the space provided below for your answers.

2) Compare your answers with those given at the end of the unit.

1. What functions do maps serve in journalism?

.....

.....

2. Why would it be wrong to use a choropleth map to show the total number of schools in each Indian state? What kind of map should you use instead?

.....

.....

16.5 A SURVEY OF POPULAR DATA MAPPING TOOLS

From maps, let us now move to learning about tools that can be used to make maps. The tool you choose will depend on your technical skill, the complexity of your map, and your budget.

16.5.1 No-Code Tools for Beginners

These tools run in your web browser, are usually free or have good free options, and don't require any programming.

- **Datawrapper:** It is a favourite in many newsrooms. It is incredibly easy to use and makes clean, interactive maps (choropleth and symbol). It guides you through the process of making a good map. You just copy and paste your data, and it does the hard work.
- **Google My Maps:**This tool is great for making simple point maps. You can easily add points, lines, and shapes, change how they look, and share your map. It is perfect for showing the locations of a series of events.

16.5.2 Low-Code Tools for Intermediate Users

These platforms give you more options for customisation and analysis.

- **Tableau Public:** This is a very powerful data visualisation tool that is also great at mapping. It lets you do more complex analysis, layer different datasets, and build interactive dashboards that include maps. The learning curve is steeper, but it is much more flexible.
- **Flourish:** It is known for making beautiful, animated data visualisations, including cool map types like 3D maps and flow maps. It is built around a story-driven approach to making presentations with maps.

16.5.3 GIS Software and Code Libraries for Advanced Users

If you want total control, need to do custom analysis, or want to create a truly unique map, you will need to use these tools.

- **QGIS:** This is free, open-source, and professional-level Geographic Information System (GIS). It is the top choice for any serious map

analysis because it lets you edit map files, run complex spatial queries, and get your data ready for other tools. It is a desktop programme with a very steep learning curve.

- **Mapbox / Leaflet.js:** These are JavaScript libraries for building your own custom interactive maps for the web. You need to know how to code, but they give you complete control over the design, interactivity, and data. This is how places like the New York Times build their well-grounded, award-winning map stories.

16.6 CRITICAL CONSIDERATIONS IN MAPPING

A map may seem like a piece of scientific evidence, but it is made by a person making choices, and every choice can introduce bias. A good journalist has to be aware of these traps, which include the following:

16.6.1 The Modifiable Areal Unit Problem (MAUP)

The story your data tells can change depending on how you draw your boundaries. For example, a map of wealth might look one way if you group the data by state but completely different if you group it by district. State-level averages can hide small pockets of extreme wealth or poverty. Always be aware that the geographic units you choose might be hiding part of the story.

16.6.2 The Ecological Fallacy

This is a logical error that occurs when you make inferences about an individual based on aggregate data for their group. For example, if a district has a high literacy rate, you can't assume that any random person you pick from that district is literate. Your map shows trends for areas, not for the people in them. Your language must reflect this. Rather than saying "People in this area are poor," say, "This area has a high poverty rate."

16.6.3 How Classification and Colour Choices Create Bias

The way you group your data (classification) and the colours you use can completely change the message. You can make a map look scary or harmless just by changing the class breaks in the legend.

The right thing to do here is to be transparent. Always explain in your methodology why you chose a certain classification method. Use tools that let you try different methods so you can pick the one that tells the most honest story about the data.

16.6.4 Privacy and Anonymity in Point Maps

When you are mapping very specific data, like the location of individual disease cases or crime victims, you run a serious risk of violating someone's privacy. Plotting an exact point on a map could make it possible to identify a person's home. A journalist needs to protect the identities of individuals. Either group the data into a larger area (like a neighbourhood or a census block) or "jitter" the points by adding a tiny bit of random error to their location so they don't land on the exact address.

16.7 CASE STUDY: MAPPING ELECTION TURNOUT IN INDIA

Having learnt the basics of geographic data and map-making vis-à-vis data journalism, let us now do a case study on mapping participation in a national-level election in India.

Scenario: A data journalist wants to show the voter turnout from the recent Lok Sabha election for all states and Union Territories to see where participation was highest and lowest.

Workflow in Action:

1. **Story and Question:** The question is, "How did voter turnout vary across India in the 2024 Lok Sabha election?" The goal is to make a choropleth map for a news article.

2. **Data Acquisition:**

- **Statistical Data:** The journalist gets the official turnout data from the Election Commission of India (ECI) website. It is a spreadsheet with state names and their voter turnout percentages.
- **Geographic Data:** The journalist downloads a GeoJSON file of Indian states from a reliable source like Natural Earth.

3. **Joining Data:** The journalist uses Datawrapper. They upload the GeoJSON file, then copy and paste the data from the ECI spreadsheet. Datawrapper automatically tries to match the "State Name" column from the spreadsheet to the state name property in the map file. The journalist double-checks that every state was matched correctly.

4. **Choosing Map Type:** The data is a rate (a percentage), so a choropleth map is the right choice.

5. **Designing the Map:**

- **Classification:** The journalist uses Datawrapper's default "Natural Breaks (Jenks)" setting, which is good at finding logical clusters in the data. They pick 5 classes to make the visual differences clear.
- **Colour:** They pick a single-colour sequential palette (from light blue to dark blue), where darker colours mean higher turnout. This is an intuitive choice.
- **Context:** They give the map a clear title: "Voter Turnout in the 2024 Lok Sabha Election." They add a legend that clearly shows the percentage range for each colour. They add the source: "Source: Election Commission of India."

6. **Review and Annotation:** The journalist looks over the map. They see that states like Nagaland and Manipur had very low turnout. They use Datawrapper's annotation tool to add a small note right on the map, pointing to these states and saying, "Turnout in some northeastern states was significantly lower than the national average." This gives the reader extra guidance. They also turn on tooltips, so when a user hovers over a state, they'll see the exact turnout percentage.

The final product is an accurate, ethical, and informative map that clearly tells the story of how voter participation differed across the country.

16.8 LET US SUM UP

Knowing how to map geographical data is a crucial skill for any data journalist today. This unit gives you an insight into how to put information in context, reveal spatial inequality, and tell stories that are grounded in a real sense of place. In this unit, we have walked through the whole process, from the basics of geospatial data to the final product. We have looked at what maps can do in journalism and surveyed the most common types, stressing how important it is to pick the right visual for your data and your story.

We also waded into the ethical minefield that every journalist has to navigate when making a map. Understanding how your choices about classification, colour, and scale can create bias is the key to producing work that is not just eye-catching but also honest and responsible. By mastering these concepts and tools, you can unlock the stories hidden in the data and present them to your audience with clarity, accuracy, and impact.

16.9 KEYWORDS

Geospatial Data: Information that has a geographic component, linking it to a specific location on Earth.

Vector Data: A type of geospatial data that represents features using points, lines, and polygons.

Choropleth Map: A map that uses differences in shading or colouring within predefined areas to indicate the average value of a property or quantity in those areas.

Normalise Data: The process of adjusting raw counts to account for differences in population or area, typically by calculating a rate or percentage. This is essential for creating accurate choropleth maps.

GeoJSON: A standard, web-friendly file format for encoding vector geographic data.

Shapefile: A common, multi-file format for storing vector geographic data.

Join (or Merge): The process of combining statistical data from a table with a geographic data file based on a common matching column (key).

Data Classification: The process of grouping data into classes or bins for visualisation on a map, for example, using methods like equal interval, quantiles, or natural breaks.

Ecological Fallacy: A logical error in the interpretation of statistical data where inferences about the nature of individuals are deduced from inferences about the group to which those individuals belong.

16.10 FURTHER READINGS

Cairo, A. (2016). The Truthful Art: Data, Charts, and Maps for Communication. New Riders.

Monmonier, M. (1996). How to Lie with Maps. University of Chicago Press.

Peterson, G. N. (2018). GIS Cartography: A Guide to Effective Map Design. CRC Press.

Online Resources:

- Datawrapper Academy: https://academy.datawrapper.de/
- QGIS Training Manual: https://docs.qgis.org/3.40/en/docs/training_manual/index.html
- The Pudding (for creative examples of visual and map-based stories): https://pudding.cool/

16.11 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress - 1

1. Vector data uses basic geometric shapes to establish clear geographical information. It uses points, lines, and polygons to show things like a hospital or a city (a point), a road (a line) or a state (a polygon). Raster data, on the other hand, uses a grid of pixels to show continuous information without sharp edges. This grid-based approach makes raster data perfect for showing things that don't have clear boundaries but instead blend into each other like temperature or a satellite photo.
2. Such a collection of files is called a Shapefile. These files have to be kept in the same folder since each of them contains information about different geographic features which is then put together and interpreted by GIS software.
3. Geographic data gives spatial information about geographical features and includes map files, the shapes of states, and the like. Statistical data refers to the numerical information which is entered into spreadsheet and then displayed on maps, such as the population of each state. Mapping is the process of combining statistical data with geographic data.

Check Your Progress - 2

1. Maps serve several important purposes in journalism. Most significantly they give information about the location of the story and how information pertaining to that story occurs on ground through that geographical area. They help us do a comparative analysis of relevant data from multiple locations and show relationships and connections between them.
2. It will be incorrect to use a choropleth map in this case because a choropleth map is used for rates or densities, not raw counts. Using it for the total number of schools would be misleading, as larger states would appear darker simply because they are bigger and have more schools, not necessarily because they have better school coverage.