# Introduction:

* GIS: Geographic Information System are frameworks which gather, manage and analyse data
* Disciplines:
  + GISystems: Focuses on tech and tools
  + GIScience: Field of research which studies the theory and concepts which underpins GIS
  + GIStudies: Study of use of Geographic info
  + GISS: Geographic Info Science and Systems
  + GIS&T: Geographic Info Science and Technology
* Uses and applications:
  + Can track objects via real-time maps
  + ensure security via monitoring
  + can be used to navigate open areas to respond to an incident
  + Is used to understand what has happened, what is happening and what will happen in an area
  + Can be used in disaster management, urban planning, environmental planning and monitoring

# Components:

* Hardware: Includes centralized computer servers, desktop computers and handhelds
* Software: Store, analyse and display geographic data. Also provides tools needed to input, manipulate and query geographic data
* Data: data which has location info. Integrates geographic info with data from other sources
* People
* Workflows: Composed of unique structured plans and business rules
* Geographic Information is made up of to 2 types of info:
  + Attribute: Describes non-location
  + Location: Describes where an object resides on the earth (x/y co-ordinates, latitude and longitude, others etc.)

# Definitions

* Vectorization: Raster 🡪Vector data
  + Adjacent same valued raster cells with same attributes are aggregated and then class boundaries are created at intersection between classes as vector lines
* Authalic: Theoretical sphere which has the same surface area as the Earth
  + Used to simplify calculations and models
* Latitude: Parallel east and west running lines around surface of the Earth. Used to measure distances to the north and south of Equator
* Longitude: North and South running lines around surface of the Earth. Used to measure distances to the East and West of Prime Meridian. They intersect at the poles
* Attribute Query: When you query features based of attribute
* Uncertainty: Situations in which digital representation is simply incomplete.
* Accuracy: How close a measurement is to the true/accepted value
* Precision: Consistency of repeated measurement
* Ecological fallacy: mistaken conclusion drawn about individuals based on findings from groups they belong
* Logical Error: When individuals mistakenly infer info about individuals from aggregate data
* Modifiable Areal Unit Problem is when results change depending on size, shape and delineation of spatial units.
* Scale affect: Inconsistencies in analyses conducted across different scales or spatial resolutions
* Zoning effect: Inconsistencies in analyses based on different zonal systems
* Datum: Frame or Reference for finding locations on the Earth based oin spheroid model
* Spheroid: Ellipsoid which approximates shape of a sphere or an ellipsoid created by rotating an ellipse about its major or minor axis
* Geoprocessing: Process which software tools use geographic data to create new data
* Projection: Systematic transformation of a curved surface to flat surface
* Developable surface: Surface which can be unfolded or unrolled into a plane/sheet without stretching, tearing or shirking
* Environment settings are optional parameters one can set before running a tool. These affect a tools result
* Intermediate data: Datasets which aren’t kept, only used during model execution and immediately deleted after
* Data Evaluation: Assessing quality of data
* Data Engineering Workflow: Processes of preparing data for analysis
* Metadata: Data of data or data which describes data
* Projection: Systematic transformation of a curved surface to flat surface
* Conformal: Shapes and angles preserved
* Equal Area: Areas are preserved but shapes are distorted

# Capabilities

* Spatial Analytics: User of tools based on spatial algorithms to identify trends, patterns and relationships in data
* Fields Operations: Access the same authoritative data sources via mobile devices
* Mapping and Visualizations: View data from multiple sources all at once
* Real-time GIS: Real-time decision making via monitoring and displaying data from sensors or devices
* Transform 2D data🡪3D Models
* Imagery and remote sensing: Tools which allow people to extract info from imagery and remotely sensed data
* Data management and collection
* Able to use spatial data from other systems
* Map patterns & relationships such as:
  + Mapping locations
  + Map quantities onto a map
  + Map what is inside the map
  + Map densities such as population densities
  + What Is nearby
  + Changes
* Visualizations are either:
  + Static = can’t interact but view
  + Dynamic = view and interact
* GIS supports both continuous fields and discrete objects
* GIS datasets have a spatial component allowing one to perform tasks such as:
  + Visualize real life info
  + Implement analytical functions (proximity analysis, buffering and movement)
  + Select and filter geographically
  + Calculate different properties
* GIS allows us to see the connection between features and their attributes via the attribute table
  + Row = Geographic feature
  + Colum = Field
  + Provides descriptive info about features

# Maps

* Maps visualize where and why events happen, how far a character travels and what a country or landscape is like, give insight to why certain events happened and where they did
* Help visualize physical world
* Imagery maps represent the area.
  + Can be made by one or more images.
  + Geological information as opposed to borders
* Reference maps: Representation of an area which show borders, boundaries and other real-life features
* Thematic maps: Maps which convey information about a single topic/theme
  + Have 1+ layers which are symbolized
* Basemaps provide the background for features on a map
  + Imagery (satellite and aerial), Topographic, National Geographic, Street basemap
* WebGIS are interactive maps
  + Can be shared via Web GIS portal
  + Benefits:
    - Usable by broad audience
    - Accessible
    - Can be used my many users simultaneously
    - Access data from various sources
    - Appliable at any scale
* Able to use attributes to draw maps or style features in the map
  + Organized collection of predefined colours, symbols and their properties, and map elements
  + These control symbols used, colour, size etc.
* Cylindrical Projection:
  + Wrapping a cylinder of paper around the Earth
  + UTM (Universal Traverse Mercator) is conformal
    - Zones start at -180o and are each 6 degrees of longitude
    - UTM formula:
* Conic Projections:
  + Wrapping a cone of paper around the earth
  + Used for midlatitude zones which have an east-west orientation
* Planar Projection:
  + Sheet of paper touching the earth
  + Azimuthal projection or zenithal projection

# Data:

* Vector Data: Represents points, polygons, lines and areas and all are represented using coordinates
  + One per point
  + Lines are polygons
  + Areas are also polygons
  + More favoured in admin data
  + To reduce error:
    - Use domains to constrain valid values for a particular field
    - Subtypes: These are a subset of features with same attributes
    - Topology: Connectedness, adjacency and proximity between features. Using this we can define spatial relationships which we want protected so not matter how much we edit feature data they are adjacent to area they belong to
    - Archive data
  + Formats:
    - Shapefile
      * Can’t store null values
      * Field names limited to 10 characters
      * Represents only point, line or polygon features
    - Layer file
      * Replaces shapefile in ArcGIS pro
      * Stores info on a layer/group of layers
      * Has pointers to data and a description of how to render a layer
    - GeoPackage:
      * Stores both raster and vector data
      * Implemented as SQLite Database container which contains data, metadata, descriptions and constraints. We can use SQL queries to access
      * Preferred format for QGIS
    - GeoJSON
      * Uses JSON (JavaScript Object Notation) files to encode geographic data and features alongside associated attributes
      * Written and maintained by non formal standards organization
* Raster data:
  + Divides space into rectangular cells
  + Objects are collections of cells
  + Properties/attributes are assigned values to cells (single val)
  + Favoured in remote sensing & elevation
  + As cell size increases, raster becomes more voluminous
  + File types:
    - JPEG
    - PNG
    - TIFF
    - BMP
    - GIF

|  |  |  |
| --- | --- | --- |
| Data Collection Methods | Raster | Vector |
| Primary.(Direct measure)  Process to collect data specifically for GIS use | Digital remote sensing images | GPS measurements |
| Digital aerial photographs | Survey Measurements |
| Secondary(Indirect measure)  Process of creating raster and vector and DBs from other files | Scanned maps | Topographic maps |
| DEM from maps | Toponymy (place name) data sets from atlases |

* Data collection (compromised of data capture and transfer) workflow:
  + Planning: User requirements, gather resources, develop plan
  + Preparation: obtain data, clean data, scan map into digital format
  + Digitizing/transfer: Needed for secondary data capture to digitize the scanned maps or transfer other formats into GIS
  + Edit/improve: Validate data, correct errors, improve quality of data
  + Evaluate: Check if what we have is qualified to be used for other tasks
* Collecting data is expensive at takes up to 85% of workload
* Buy vs build is important when thinking about data transfer
* Field data types:
  + Object identifiers. Unique id for each row in table, don’t query using these ones
  + Numbers: Can be short, long float or double. Usually queried via quantitative values
  + Text. Queried via textual values
  + Dates. Queried via time frame
* Attribute data can be entered via:
  + Manually
  + Direct data logging
  + Text/voice recognition
  + Citizen Science
  + Questionnaires
* Types of vector spatial analysis are:
  + Queries (Location and attribute)
  + Measurements
  + Transformations
    - Clip
    - Cut
    - Buffering
    - Intersect
    - Union
  + Descriptive summaries such as Sum. Min Max etc
  + Optimization
* Types of Raster Analysis:
  + Reclassification:
    - Binary masking: Assign 0 or 1 to each cell in input using lookup table
    - classification reduction: Assign new values to classes/ranges to reduce no of classes
    - Classification ranking: Assign ranks to unique categories/values
    - Changing measurement scale: Assign vals to a qualitative scale layer to generate quantitative layer
  + Overlay: Creating a new layer by assigning each cell value which is a function of the independent values of the cells at the same location on 2+ input layers
    - New = Layer1 OP Layer 2 where OP is either a logical or arithmetic operation
  + Zonal Statistics: Summarize info about 1+ features within a zone
    - Dataset MUST be raster
  + Cost Path Analysis
    - Least Cost Path(LCP): Most cost effective path finding method
    - Practice in linear infrastructure and route finding apps
  + Terrain Analysis
    - Using digital elevation data(DEM) to analyze landscape topography
  + Interpolation
    - Estimate unknown values on existing sample points

# Representation of Data

* Scale on map is expressed as either:
  + Scale bar, akin to a ruler
  + A ratio
  + In two equivalent values
* Features: Display objects on a map using GIS
* Digital data is easy to transform, process and analyse
* Benefits of Digital Representations:
  + Easily copied and transmitted
  + Simple
  + Reliable
  + Economies of scale
* Issues:
  + Rarely perfect as either details are irrelevant, expensive or voluminous
  + Uncertainty: Important to know what’s missing
  + Geographic world infinitely complex
* Ways to represent geographic information:
  + Discrete objects:
    - World as empty space
    - Objects have well defined boundaries
  + Continuous fields:
    - Continuous source
    - Represented by some variable/attribute which is measurable at any point
* Ways to represent fields (All reduce variation of field to set of objects and attributes):
  + Regularly or irregularly spaced sample points
  + Rectangular cells
  + Irregularly shaped polygons or network of triangles
  + Polylines which represent contours
* In geographic coordinate systems, Earth shape uses three things:
  + Datum
  + Prime Meridian
  + Unit of measure
* To list, we first include latitude then longitude i.e. 0oN , 78oW
  + 0 Degrees north of equator and 78 Degrees west of Prime meridian
  + North and West are +, South and East are -
  + Here we use degrees/minutes/seconds referred to as DMS
  + Decimal Degrees (DD) = DMS but minutes/seconds are expressed as decimals
    - Used as storage/computations are faster
  + DMS🡪DD:
    - Treat 1 degree as 1 Hour. Using this to convert seconds🡪degree, div by 3600 and for a minute, divide by 60. Afterwards add up all degree values to get values in DD
  + DD🡪DMS
    - Multiply Decimal by 60 to get minutes part, then split the output into numbers pre and post decimal. The pre decimal is minutes. Multiply the numbers post decimal by 60 to get second
* Projection Coordinate system tells data how to draw on a flat surface while a Geographic coordinate system is unprojected
* Georeferencing: mechanism which allows you to use geographic information to identify locations on a map
* Types of georeferencing include:
  + Using XY coordinates
  + Place names
  + Postal codes/addresses
  + Street address
  + Grid system
* Spatial Analysis: Process of looking at and analysing a map by identifying patterns, trends or making decisions
  + How people understand the world by turning data into info
* 6 Spatial analysis categories:
  + Understand where you are or what is around you
  + Measure size shape and distribution like area, length, perimeter, height or volume
  + Determine how places are related: what is near, within or how something overlaps in space and time
  + Find the best locations and paths:
  + Detect and quantify patterns looks for patterns, hotspots or outliers and how they change over time
  + Make predictions
* Spatial analysis workflow: Ask questions🡪Explore and prepare data🡪Analyse and model🡪Interpret Results🡪Repeat and Modify🡪Present results🡪Make decisions
* Spatial analysis requires both attributes and locations of objects
* Geography Inquiry Process (can apply anywhere with geographic components):
  + Ask Geographic questions
  + Acquire geographic resources
  + Examine geographic data
  + Analyze geographic data
  + Act on geographic knowledge
* A geoprocessing tool usually send an input layer through an analysis tool which outputs a result layer

# Models

* Definition: Abstraction of Reality used to represent an object, process or event
* ModelBuilder is a built in application in ArcGIS
* Reasons to create a model in ModelBuilder:
  + To see visual representation of analysis and geoprocessing operations
  + Automate and manage workflows
  + Run complex succession of processes as one tool
  + Plug in extra tools and parameters
  + Share geoprocessing workflows and models with other users
* Geoprocessing models automate and document spatial analysis and data management processes
* They automate workflows by adding data and tool elements to model and connecting them to form a workflow
* ModelBuilder allows one to:
  + Build a model by adding and connecting data tools
  + Iteratively process every feature, class, raster, file or table in workspace
  + Visualize workflow sequence as an easy to understand diagram
  + Run a model step by step or run entire model
  + Make ones model into a geoprocessing tool
* Elements of a model:
  + Variables: Data or variable values
    - Displayed as ovals (green = Derived/output data, blue is input data, input value is green-blue while derived value is light blue)
  + Tools such as script/model tools
  + Groups are visual categories which include other elements in model. Can assemble processes into logical units
  + Connecters
    - Denoted by Lines
    - Precondition connects one process to another to show the previous one must be done before the next one
    - Feedback connecter is used to feed the output of the model back into the first process and modify it again a set number of times
  + A diagram of a company

    Description automatically generated
* States of a model are:
  + Not ready to run: Denoted Grey
  + Ready to run: Displayed in Colour
  + Already run: Coloured and have a shadow/shading
* Model environmental settings

|  |  |  |
| --- | --- | --- |
| Setting type | Description | Override Characteristics |
| Application level | Set in Environments: Applied to all tools | Becomes the default setting after being set; is applied to any tool executed |
| Tool Level | Set in each individual tool dialogue box; specific tool only | Applied to a single run of a tool and overrides application level settings |
| Model Level | Set in model properties; pertains to tools within model | Overrides tool-level and application level settings |
| Model Process Level | Set in each model process | Overrides all levels |

* Questions to ask before building a model:
  + Goal?
  + What data is needed
  + Most effective workflow
* Guidelines:
  + Determine scenario and criteria
  + Set goals for model
  + Get necessary data
  + Choose tools
  + Build and run a model
  + Refine results of model
* Always pass a parameter for tools and models
* Validate any model before using to see if data inputs need to be fixed or not
* Model Parameter: type of parameter exposed in a geoprocessing model that appears in a model's dialog box and allows for input
  + Can make each parameter for model into a model parameter instead of for each tool
  + Allows one to run model from tool dialog box
  + Important to set if one intends to share model
* Tool parameters determine inputs, outputs, and other values for a tool, whereas model parameters make models dynamic when run from the tool dialog box and also allow for user input

# Databases

* Database: Organised collection of information
  + All data is stored in a single, large, inflexible table
* A DBMS (Database management System): Software used to manage, access and maintain databases
  + Acts as a bridge between databases and devs, users and APPs
* Geospatial DBMS include geographic information about data points to help answer spatially explicit questions
* Different types of DBMS:
  + Hierarchical
  + Network
  + Object-oriented – OODBMS (GIS Data storage)
  + Object-relational - ORDBMS
  + Relational (ArcGIS file geodatabase is an example)
    - Collection of tables which are connected in a way that data can be accessed without reorganising the tables
    - Primary Key: Attribute which uniquely identifies record
    - Foreign key: When a primary key corresponds to another attribute in another table
    - Attribute: Column/Field
    - Domain: Type of Values stored in attribute
    - Tuple: Row/Record
    - Relation is table itself
    - Cardinality of relation: how tables are linked
    - Pros:
      * More focused
      * Separately preparable
      * Less need to edit same thing multiple times
      * Can be maintained separately
      * Good flexibility
    - Potential for redundancy
    - Join: Tables joined via common row/value or key
    - Relate: Defines a relationship between 2 tables via common field, doesn’t append attributes of one to other
* Forms of a RDBMS:
  + First Normal Form
    - No sequence to ordering of rows or columns
    - Each row is unique
    - Only contain one value
    - All values in column pertain to same subject
  + Second Normal Form
    - All non-primary keys must depend on primary key while the primary key remains independent of all non primary keys

# Geodatabase (GBD)

* Definition: a database or file structure used primarily to store, query, and manipulate spatial data
  + Native data storage format for ArcGIS
  + Able to import different datasets from various sources and use them in GIS analysis
  + Components:
    - Feature Datasets: Collection of spatially related feature classes
      * Collection of geographic features with same geometry type, attributes and spatial reference.
      * Reason of creation: Need a GDB with feature classes in feature dataset (topology or geometric network)
      * Main purpose: Store spatially related feature classes in common dataset
      * Specify spatial reference and to add feature classes of any type, must have same spatial reference as feature dataset such as same coordinate system
    - Mosaic Datasets: Collection of raster datasets/images
      * Can store manage view and query these
      * Source rasters can be combined to form mosaic dataset and dataset can display it as if they were merged
      * This is displayed as a special group layer of 3 layers: Boundary layer (displays boundary of mosaic dataset), Footprint layer (displays footprint of each raster within mosaic dataset) and the image layer (controls rending of mosaic image)
    - Standalone feature classes
    - Nonspatial table: Contains attributes
  + Records = rows
  + Fields = columns
* Types:
  + File geodatabase: Collection of GIS datasets stress in a file system folder
    - Good for users due to increased capacity and editing capabilities
    - Works across OS and can store individual datasets up to 1TB
  + Enterprise geodatabase:
    - Require DBMSs
    - Unlimited editors
    - Supports versioning and replication
    - Used in larger organizations
* Advantages of geodatabase:
  + Centralized repo: all data stored in same database
  + Scalable data model: Able to migrate data from one gdb to another upgraded format
  + Increased data integrity
  + Imagery Support by managing multiple images as one
* Geodatabase workflow:
  + Evaluate data🡪Create geodatabse🡪Organize data🡪Add data
* Schema of geodatabase defines physical structure alongside its rules, relationships and properties within dataset.
  + Can also model GIS data in an organized manner, talk about structure of any component like feature classes and tables
* After creating gdb, can share with others by exporting it as an XML

# Machine Learning

* ML is a subsection of AI
* ML is the aim of getting computers to act and learn like humans do by feeding data and information
* ArcGIS makes uses of this via geoprocessing tools and ModelBuilder
* Features in ML = Attributes in GIS
* ML Workflow:
  + Data prep🡪Data exploration🡪Model Development🡪Model Deployment
* Feature Engineering: Cleaning and preparing data for models
  + Dealing with missing values, outliers, binning log transformation, one hot encoding or normalizing data
  + Dimensional Reduction: Reduce no of attributes/columns via methods like PCA
  + Normalization is done via methods like Gaussian, Min Max scaler, Log etc
  + Imputation: Fill missing values
  + Enrichment: Add relevant columns
* Spatial Feature Engineering: Process of developing extra info from raw data using geographical knowledge
* ML uses pre-existing data exploration like charts, graphs and data tables
* ArcGIS uses both ML data exploration and spatial exploration methods
* Data exploration:
  + Clustering: Find groupings in data
  + Feature importance: Sort data by influence or importance
  + Hypothesis testing
* Non-Spatial Models:
  + Linear models
  + Random forests
  + Vector classification
  + K-Means Learning
  + Deep Learning
* Spatial Models:
  + Generalized linear models
  + Weighted regression
  + Empirical Bayesian
  + Imagery classification
  + Clustering
* Spatial ML: When spatial data, methods or algorithms are applied to ML process
* Random forest: Algorithm which combines output of multiple decision trees to reach a single result
  + Parameters include:
    - Max Features
    - N\_estimators: No of trees you want to build before taking max voting or avg of predictions
    - Max\_depth
    - Min\_samples\_split: Min number of samples required to split node
    - Min\_sample\_leaf: minimum number of samples in leaf node

# ArcGIS Specific Information

* ArcGIS Online is a cloud-based Web GIS deployment for mapping and analysis which can be accessed anywhere and at anytime
* ArcGIS Pro is an application for which has the tools for data management, visualization and analysis of 2D and 3D data
  + Cantered around projects
  + Stores multiple items, such as maps, layouts, tables and charts.
  + Key components of ArcGIS are: Ribbon Tab, group, Tool/button, view and Pane
* The contents pane displays contents of active view
* Layouts and maps are stored in a spatial analyst project
* ArcGIS allows us to use AI to recognize patterns and to create 2D/3D maps for visualization
* If we want to see a specific layer of a map in ArcGIS, we navigate to the contents pane, right click the layer then choose Zoom to Layer. To share this layer, we can do the same steps but select share
* If we want to select a subset of data in the Data Engineering view, we right click a statistic in the statistics table
* To calculate statistics in ArcGIS: Open Data engineering view, add fields to statistics panel, select the field and then calculate statistics
* Charts help reveal characteristics of data
* Data clocks are good for temporal and cyclical patterns
* We can split charts by fields to have a box plot for every unique value in a text field
* We can either delete or replace missing values to handle null
* Spatial Join Tool: This is used when a dataset needs to be spatially aggregated to support analysis goal
* Construct Data Engineering: Constructs new field from existing ones
* Always look for reputable sources when searching for a dataset
* GIS uses x,y coordinates
* Common editing tasks using spatial data include creation, deleting and modifying a feature
* ArcGIS provides: Imagery, Demographic data, data enhancement and basemaps as data products and services
* Structuring Analysis: Map spreadsheet data🡪Update metadata🡪Enhance layer with demo data
* Sources for data to use in ArcGIS include: Spreadsheets, databases (oracle and business intelligence)
* Tools which van be added to a model are system tools, other models and python script tools
* Can load multiple data sources at a time when using Feature class to geodatabase tool
* A multiuser database is good when wanting to give multiple people the ability to view and edit data in a gdb.
* To connect to data, you first connect to the folder the data is connected in



1. Basemap
2. Add data
3. Select by rectangle
4. Select by attribute
5. Select by location
6. Measure
7. Full extent
8. Fixed zoom in
9. Fixed zoom out
10. Zoom selection
11. Go to XY