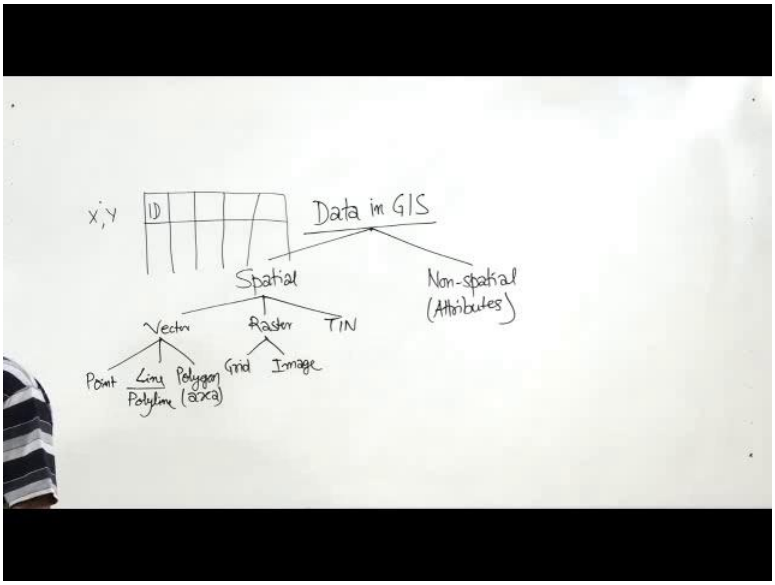
Introduction to Geographic Information Systems
Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture - 06
Non-spatial data (attributes) and their types

Hello everyone, and this is a regarding the Non Spatial Data. Far we have discussed that data in GIS can be divided in two main categories.

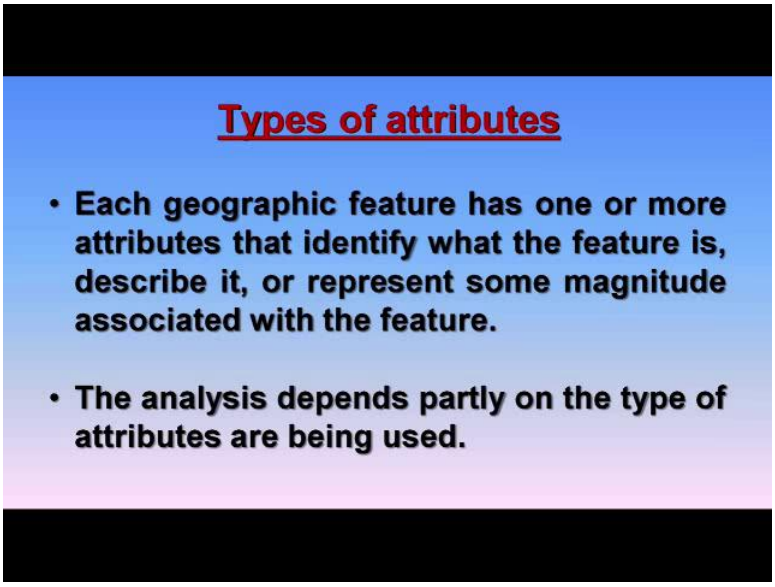
(Refer Slide Time: 00:35)



One is this spatial one, and another one is non spatial one. The non spatial one which we will discuss in detail today and we have already covered vector data, different types of vector data, raster data TIN data we have also discussed the comparison between vector data, raster data we have also gone through the comparison between raster and TIN. Now let say a focus on in the attribute and which is non spatial data as you know that with each vector data whether point line polygon theoretically we can store n number of attributes.

So, first as you know that this is also you know very simple language we say as tabular data. Because it is stored in a form of table in GIS or in GIS database as well. So, we call them as tabular data.

(Refer Slide Time: 01:34)



Now each a geographic feature whether it is a point line or polygon or vector entity I can have more than one or more than one attribute. And basically the purpose of keeping this to identify that particular feature not only identifies; that means, not only keeping you say unique id inside the database inside our tables. But also the discussion of the data if we take an example of a point data then, a point data will have spatial information that is the x, y coordinates. But in devil form it will have it can have n number of attributes. So, starting from it is id maybe is ground dimensions, and if I take example of you say ground water well then, it will have you know the who owns the well what is the depth of the well and what are the different water levels during monsoon, pre monsoon, post monsoon. And if I am having information about the water quality of that particular, well I can store in other columns as well.

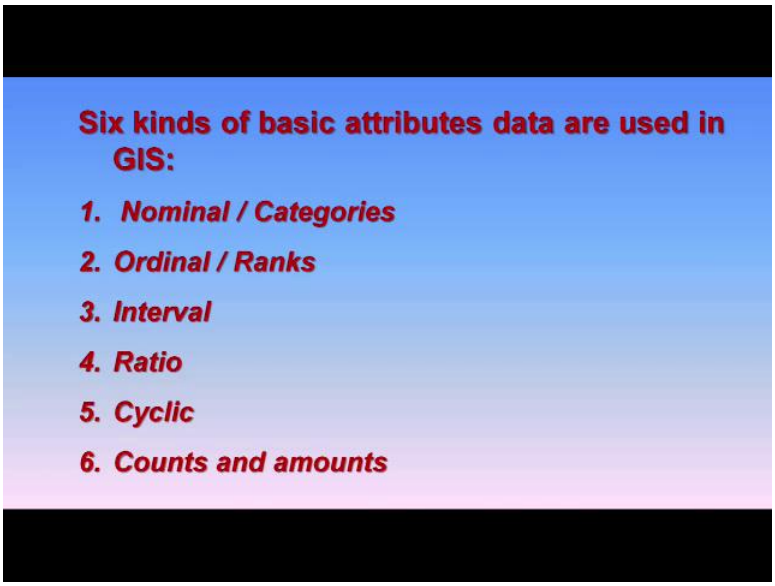
So, a basically to describe not only identify that feature that what it is and also other information's about that particular feature or some magnitude maybe the chemical qualities. Maybe if I am dealing with earthquake epicenters then spatial information that

x, y coordinates are here each I can assign id then occurrence of the earthquake maybe, the date maybe the magnitude maybe the depth of focal depth and so on so forth, all kind of details I can get associated with a point data.

Similarly, we can also do for line as well as polygon data and this more rich our attributes are more information about each vector entity when we are having then, we can later on in the analysis. We can really take lot of advantages. So, theoretically our GIS systems or support n number of columns or from database point of view n number of fields against each such objects and then, in analysis part we can take advantage of remember this that in case of raster data, we can have only one single attribute, but in case of vector data theoretically we have n number of attributes.

Now, basically if we start looking different types of attributes like, we have seen different types of vectors different types of raster's. Of course, in case of TIN, there is no other types of TIN are there the single type of TIN, but in case of attributes basics 6, kinds of or 6 types of attributes so, far have been implemented in GIS.

(Refer Slide Time: 04:38)

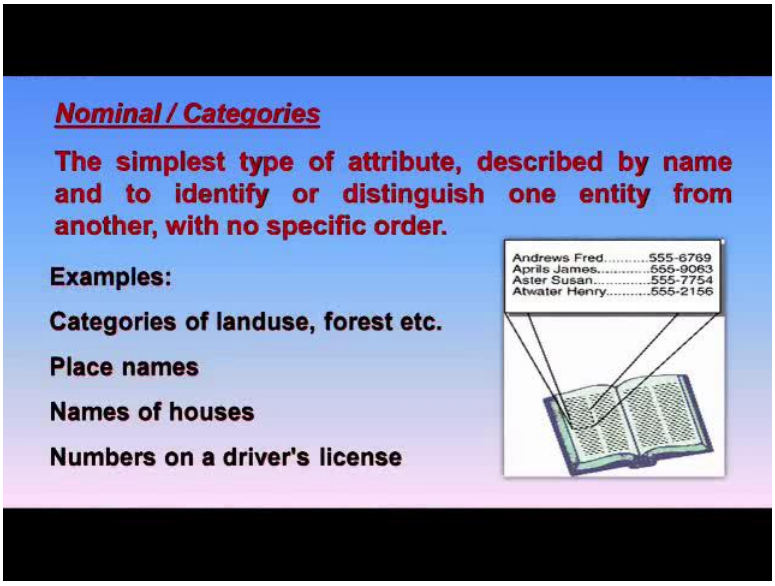


Maybe in a literature you may find some little different numbers, but if you look carefully you would that we can maximum categorize into 6 categories. So, let us go 1

by 1. The first one in this category is the nominal or we also called as category. Then another one is the ordinal and ranks we will see in detail about each of such type of attributes then interval then ratio cyclic and counts and amounts. So, these 6 types of basic attributes existed for have been implement. So, far into GIS if new development takes place if, new type of data start coming then probably in future we will see one more type or few more types of attributes. But a whatever; so, far different types of attributes which can come into GIS which we handle today all can be handled using either one or two types of these attributes.

So, let us look the first one, which is the nominal attribute. Nominal attribute like in English grammar, we say you know the proper noun you say name and there may not be any you know sequence or order. So, there is a simplest type of attribute.

(Refer Slide Time: 06:04)



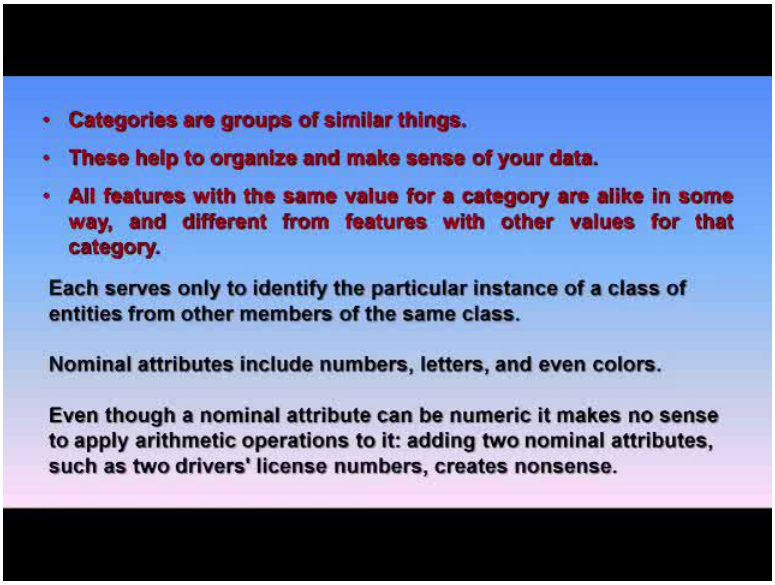
Described by it is name and the purpose here is to identify that object using that name or distinguish one entity from another and no it is specific order. So, this is underlined part here is that in nominal attributes no specific order though, you may arrange say names in alphabetically order or in seniority order or some order, but from attributes points of view from GIS point of view these attributes nominal attributes do not have any orders.

If you are using already using like excel or some other database or GIS you would find when you start putting the data in attribute. And if you do not define the type of data which is going to come in each column then, normally it will go as a nominal data by default; that means it does not have any order and once it does not have any order. It simple name and that mean also that lot of arithmetic operations cannot be performed on nominal types of data, but sometimes we have to use because this is easy and easy to understand, easy to identify. For examples maybe you might be using a map in which you are keeping different types of categories of land use maybe soil maybe forest or geological names.

Places names again these are proper noun, there is known if you even if you order them, it does not have any meaning that two adjacent alphabetically ordered places will be geographically will be located adjacently it is not true. So, the places names of houses and also even if you use numbers and put them in nominal though their numbers you might feel that I can do some arithmetic operations. Even then if you perform arithmetic operations on say drivers' license number like if I had two driver license number and the third one will there the output for the value which will come through the addition of to drive driving license number will not carry any meaning. So, nominal attributes are just simple to describe or identify certain properties certain features on the map or distinguish them from one entity to another without any specific order.

Now, the next one comes in categories are groups for similar things these one has to remember that these are grouped together of similar things.

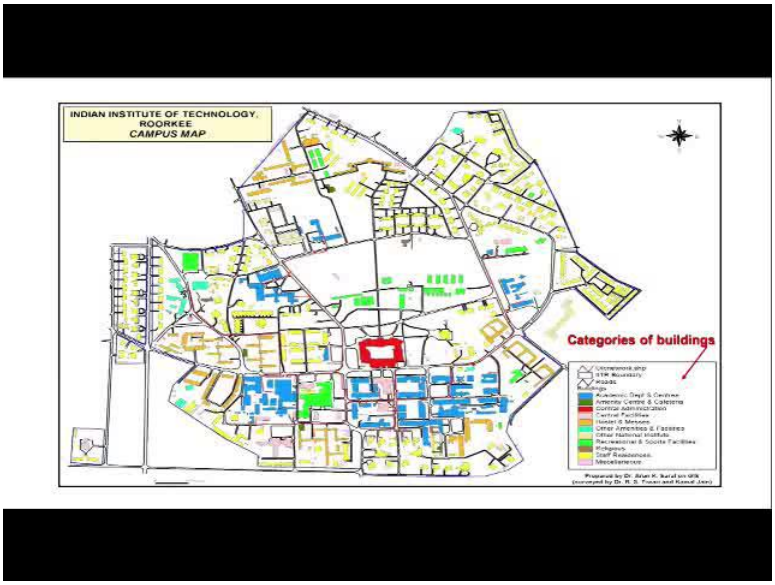
(Refer Slide Time: 08:48)



These help to organize and make sense of your data all features with same value for a category are liked in the same way and different from and different features and each serves only to identify the main purpose here, is to identify. There is a like a proper noun in typical English grammar and nominal attributes include, can include numbers can include letters both or even colors or pattern, but they will not have any meaning here and even though a nominal attribute can be numeric as I have given already example there it make no sense to apply arithmetic operations to it. For example, adding two nominal attribute such as two drivers license number creates nonsense.

An example I have taken from a map of IIT Roorkee, which was prepared on GIS platform and I have categorized different buildings, in different categories though I have ordered them.

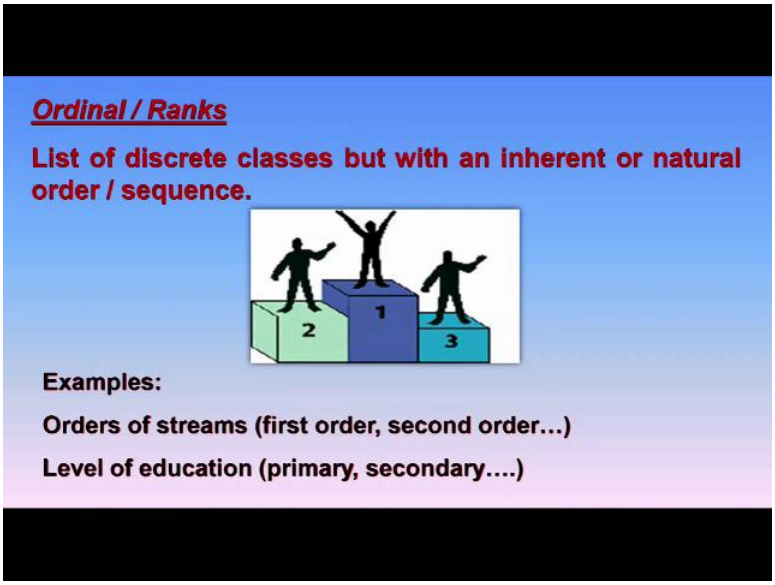
(Refer Slide Time: 09:45)



Here in alphabetical order that does not mean that geographically we have got any order or nothing for example, if this is the color scheme which be a followed is an internationally followed color scheme by the architectures. So, the red color is showing the central administrations, main administration, residential staff, residential are shown in yellow color and so on so forth. So, this is how a map when you apply or when you use the buildings as I category or nominal attribute. Then you can and display like this you can assign any color, but a we if we follow some standard color scheme then, this is how it may turn now we will see the example then instead of using a nominal attribute when we change the attribute to another type how these buildings or they their appearance will come differently.

So, the second one is the ordinal or ranks here name applies it must be having order. So, list of discrete classes.

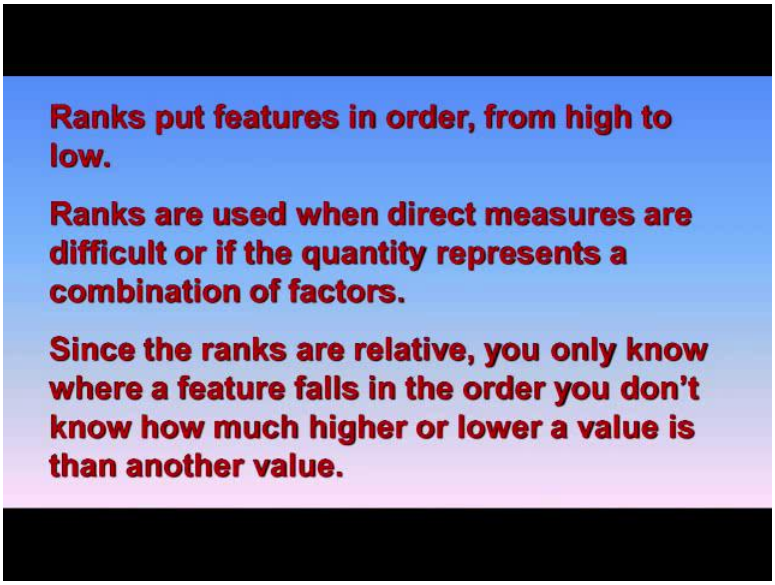
(Refer Slide Time: 10:59)



But with an inherent or natural order or sequence here this diagram also shown here in this schematic that you might be having first ranking second ranking, third ranking; that means, there is a order and there is a meaning between first and second. There is a meaning between first and third as well. So, the examples in real world in our in like in sciences we see the order of streams maybe first order second order third order and then, we say when we say third order streams we know that what is the status of third order; that means, there two more higher order streams exist and so on so forth. So; that means, there is inherent order or natural sequence exist with the data.

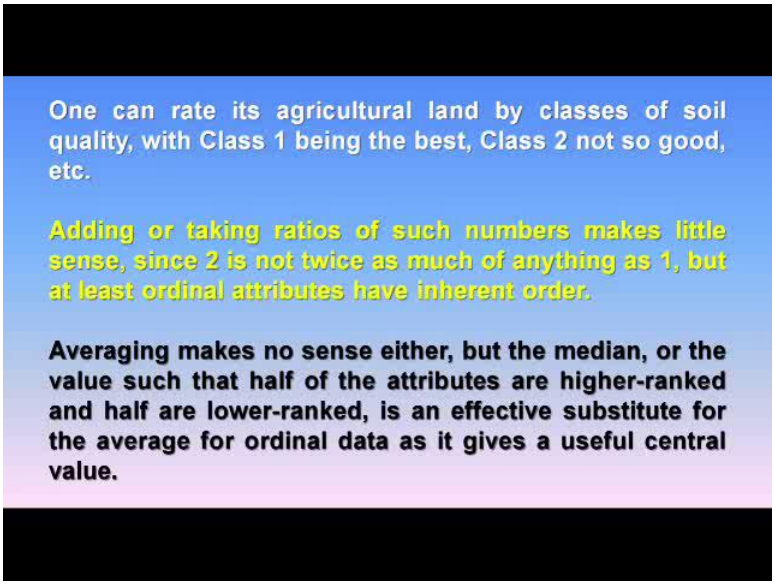
In case of nominal no order and another example can be level of education like primary secondary undergraduate postgraduate doctoral post doctoral. So, all these once we say somebody are postgraduate; that means he has done under graduate. So, there is a order and he is following that order is a kind of stairs is sequence is there and the ranks put features in order from higher to lower it depends on your requirements sometimes you also put lower to higher.

(Refer Slide Time: 12:18)



Ranks are used when direct measures are difficult or if the quantity represents a combination of factors and since the ranks are relative because, it has got a relationship the first rank and second rank has got some meaning. So, you only know where the feature falls in the order you do not know how much higher or lower a value is than another value. Because if I have already mentioned the highest one then I do not know what other higher and things are there. Suppose, if I say the person has done graduation, but if I do not have information, whether he has done post graduation or doctoral then that that is the only information which I am having at that stage anyway. So, one can rate this maybe in the agricultural land different types of soils a class one which are the best class two not so good and so on so forth.

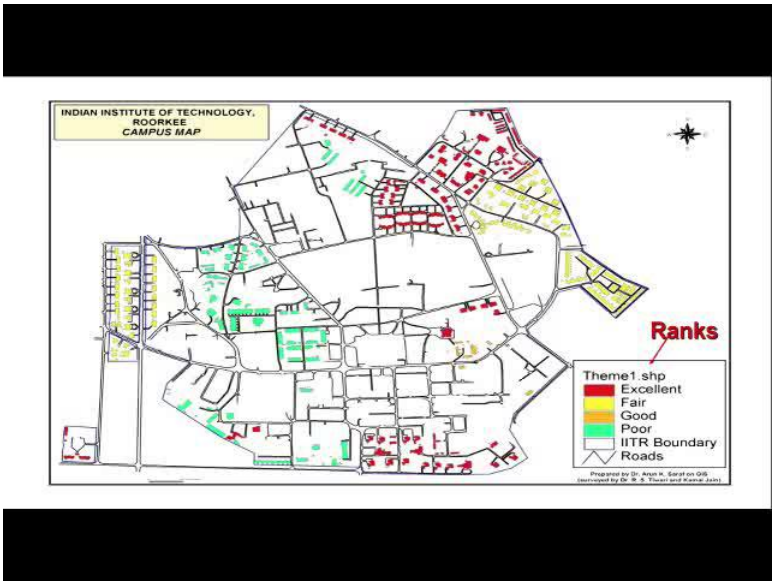
(Refer Slide Time: 13:09)



So, you can rank them and adding or taking ratios of such numbers makes little sense there you can represent ranks three numbers, but again performing arithmetic operations may not bring any results. For examples since not two is not twice as much as anything as one, but at least ordinal attributes have inherent order not like nominal attributes which do not have any order. Now everything makes no sense as I have already mentioned that arithmetic operations will not make any sense. So, far neither in nominal attributes nor in ordinal attributes, but the median or other values such as the half of the attributes are higher ranked or higher lower ranked is effective substitute for average for ordinal data. So, it has good very limited use in case of if you want to perform arithmetic operations.

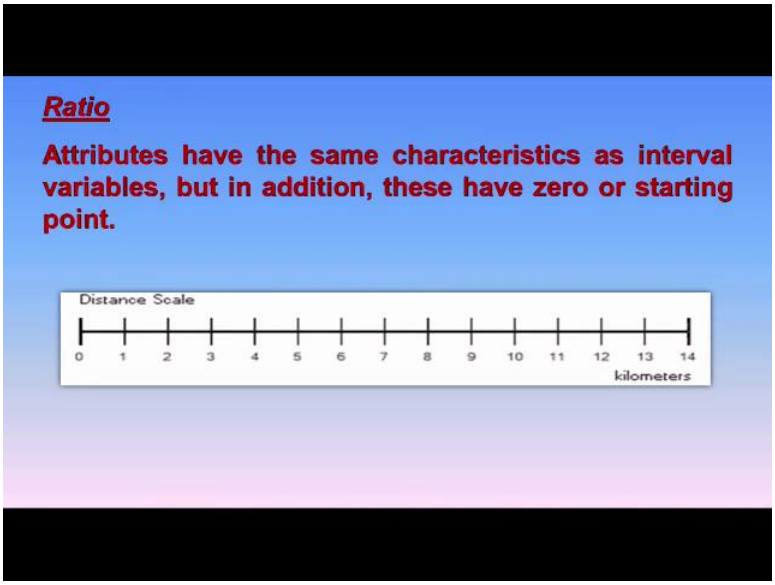
Now, let us take the same example of IIT, Roorkee map and I have ranked the buildings.

(Refer Slide Time: 14:19)



Based on their construction quality just putting them, you know may not be in real sense this it is reflecting anything, but just for representation and what we are seeing that there are certain buildings, which are I have ranked them as excellent some are having fair good and. So, there is now ranking there is some inherent order in the representation.

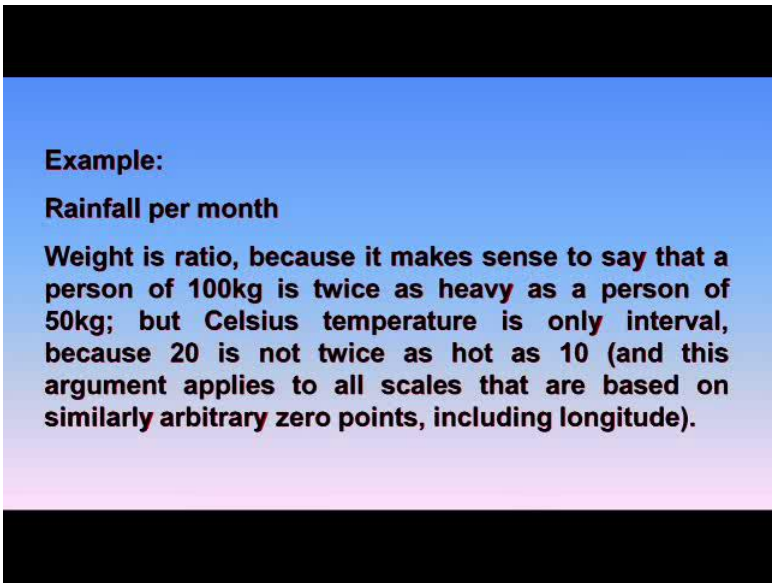
(Refer Slide Time: 14:44)



Now, the third attribute is interval. So, as we go higher and higher attributes there will be some addition in the condition or a about the understanding of these attributes like a in case of nominal there were no sequence or order or a began or end was there. But here now in case of attributes, which are interval attributes have natural sequence. So, one higher degree in case of ordinal we had some order here through is a natural sequence what in addition. Now one more thing has been added here that the distance between values have meaning and example is like a scale of Celsius temperature is interval and interval because it makes sense to say 30 and 20 degree are the same having the same difference as the twenty and ten there is a difference of 10.

So, there is a the interval attribute will have natural sequence one and in addition it will have the distance between values will have the meaning and example is also given here for the Ph scale as well that is also an interval scale now the forth type of attribute is a ratio attributes and that have the same characteristics as interval variables, but in addition. So, all those two conditions inherent order and that the difference between and distance between two values, have the meaning plus in addition there is a it will have a 0 or a starting point. So, like here a in a scale for the representing some distance. it is having a 0 and this is how we generally in the maps we put a scale bar we put a 0 and some value on the right hand side. So, likewise it will have a order it is having a sequence plus it is having a meaning like a 2 minus 1 is 1 kilometer same would be the three minus 2 equal to one kilometer and in addition it will have a starting point at 0.

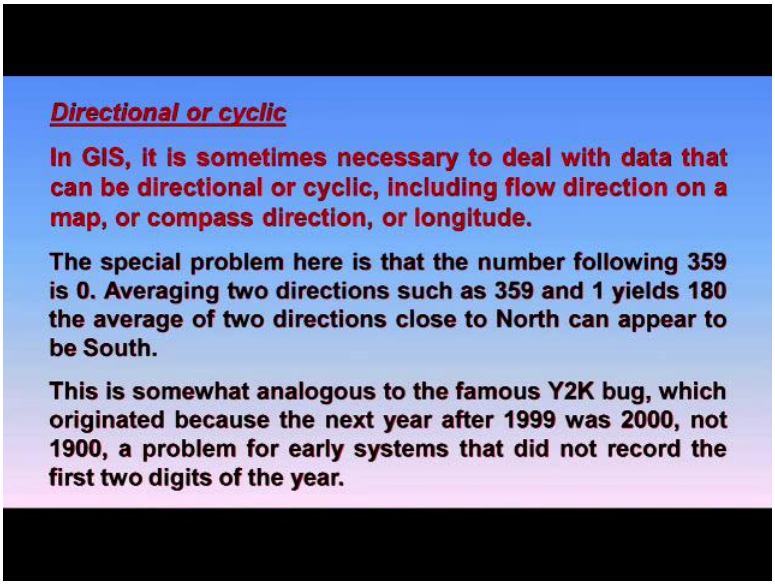
(Refer Slide Time: 17:13)



For example; rain fall per month. So, rainfall and the weight or weight is a ratio, because it makes sense to say that a person of 100 kg is twice as heavy as a person of 50 kg. So, there is a 0 value there is a starting point plus all those things which are there in the ratio are all here and the Celsius temperature is only interval because 20 is not twice as the 10 and these argument applies to all the scales that are based on similarly arbitrary 0 points including latitude longitude which we uses in our GIS maps.

Now, the fifth attribute in this category is directional or cyclic this is very, very important in GIS because the latitude longitude which would which we are using is a cyclic same like time is also a cyclic. Therefore, the handling though they are these are also numeric values, but their handling between GIS is very, very important very carefully these should be handled we will see the examples here that a sometimes it is necessary to deal with the data that can be directional or cyclic.

(Refer Slide Time: 18:25)



One more type of data in geological sciences or in we use for directions a bearing. So, bearing is also a directional type data it has to be handled differently and it has to be declared before, we enter into the system properly it should be declared that now data which is going to come in this particular column or field in my database is directional data because their normal arithmetic operations cannot be performed. We will see the example that like including flow direction on a map or compass direction or longitude. So, all will complete one cycle may be at three 60 degree may be at ninety degree maybe like time it will complete in at 60 seconds and then it becomes one minutes. So, it is a cyclic data.

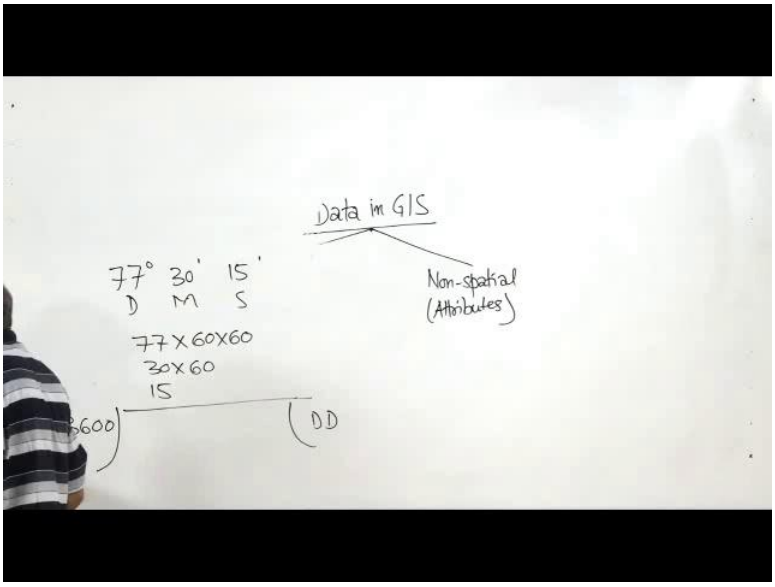
Let us take this example of that the special problem that when three 59 and it is 0 after 359, 0 comes not 360 degree and in case of degrees and average in two directions such as 359 and plus one is 180 and there would be the just opposite direction so; that means, on directional data simple arithmetic operations cannot be performed otherwise, you will go instead of north, you may lead towards the south and it will give you completely opposite directions. So, it has to be very carefully handled directional data in GIS all modern GIS softwares are nowadays capable to handle directional data very easily. But you have to declare to that system and accordingly you have to handle the data and there were when year 2000 year about to come in the year 99 people realize before that there

will be some problem because in that time the operating system used to store for year only last digits.

So, after 99 if we would not have changed our operating system then after ninety nine that double 0 would have come for the year., now double 0 maybe meaning of 1900 maybe 1800 maybe 2000. So, later on this problem by solved by instead of keeping 2 digits, two last digits for the year these start age is storing four digits for the year and then this solved problem are solved, but this problem is very famous problem of that time it is known as the y two k bug and a because of and this reason.

Another set of problems will arise because a in GIS instead of using degree minutes and second data if we convert this data into d d and once the d d is converted then arithmetic operations can be performed easily now before that we have to be because this is cyclic data. So, 60 seconds will be a will become your one minute and 60 minutes will become one degree, this instead of having a 0 to 60 scale. If we rescale 0 to 100 then it becomes d d and then, d there is a degree decimal data is falling in decimal system and then all kinds of arithmetic operations can be performed. So, in GIS when we use the geographic data in geographic projection we use instead of d m s we use d d and it is quite easy to convert from d m s.

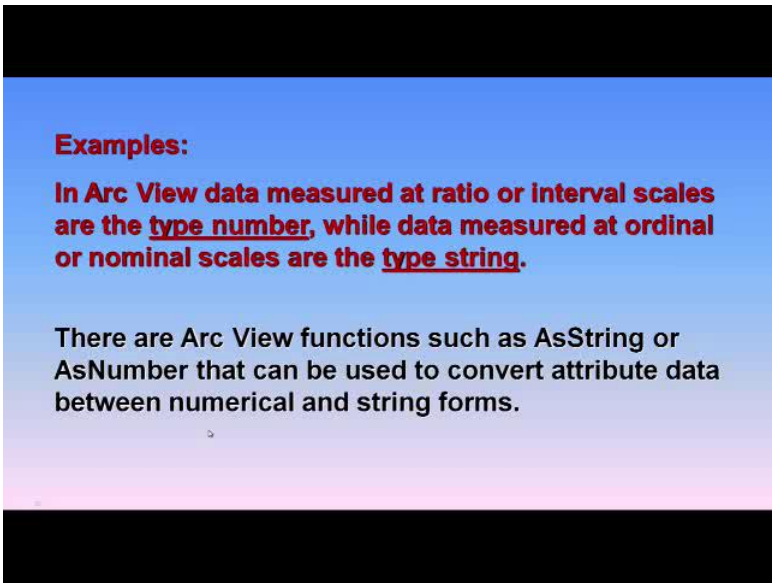
(Refer Slide Time: 22:16)



That means, degree minute seconds to d d by just a applying simple concept and calculations like, if I am having a say 77 degree, 30 minutes and 15 seconds and I want to convert this is d m s format I want to convert in d d then, what you have to do you multiply by 60 multiply by 60 because, we have to bring everything to the second that is our you know the least count or the unit here. Then 30 multiply by 60 and then you add 15, whatever you will get you divide that one by 3600 and then you will get here this d d and then arithmetic operations can be performed easily.

So, in modern days DIS are capable of converting from d m s to d d and vice-versa. Otherwise if not then simple tools can be used simple concept can be used to convert them and put them, in d d. So, later on you can change it or keep it and then even arithmetic operations it can also be performed like in example one of the earlier very popular GIS softwares.

(Refer Slide Time: 23:27)



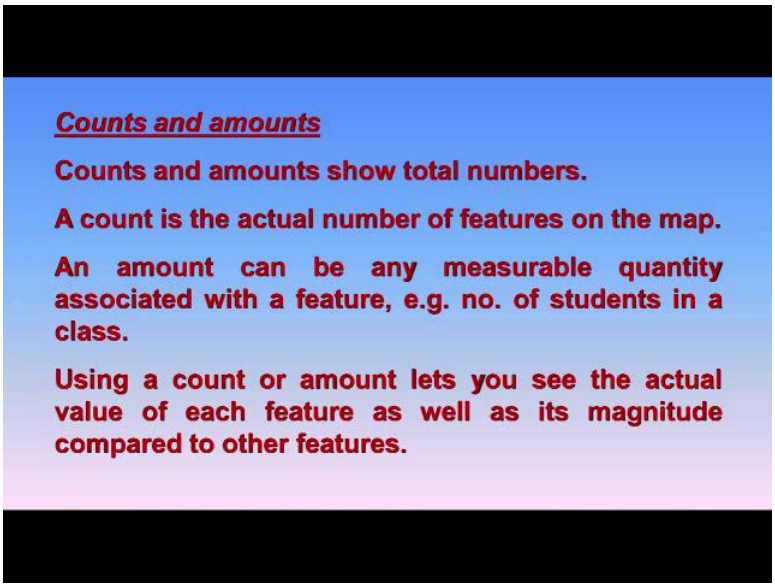
In r GIS which is modern days very popular GIS software the data measured in ratio interval scales are type number. That means, that when we put in the attribute table you put them as number as numeric value that is why the word type number is used while the data measured at ordinal or nominal scales are type string; that means, you are declaring to the system that my data can come in as strings that maybe have alphabets or alpha

numeric. So, this is how you declare in case of before entering the data and a in these functions are used in this particular software to as a string or as number that can be used to convert attribute data between numeric and string forms.

Now, one thing one has to be remember that a when you are reading your own GIS database different columns or different fields in your database their value their properties has to have to be declared very properly. Even if you are using excel and if you know that my data is going to be cyclic data declare as that it is going to be the date. So, then later on you will not have problems. But if you do not declare then by default it will take as a nominal data and then later on if you convert that nominal data to interval or ratio you might have lot of problems. So, it is better before entering the data if it is possible properly think that what kind of data is going to come in that particular field and declare the properties accordingly.

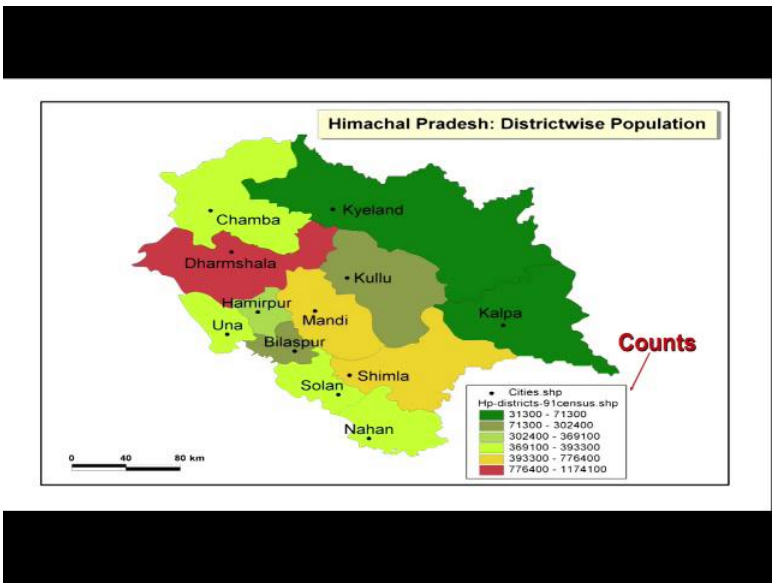
One more thing is that if you are bringing the decimal data and that too is the real numbers then you have to be also careful by declaring the number of places after decimal. If you know that your data is not going to come up to fourth places of decimal then do not put in the format for play because unnecessary it creates the redundancy in the data and later on during your analysis it might give you lot of problems. And because if you reduce later on it may bring some problems or if you increase the number of places again it might bring problem. So, better to understand that first what kind of data is going to come in this particular field and accordingly you must defined, and define the definition or properties of that field including your decimal places. So, later on you do not encounter any problem like rounding of problems or other thing, even if you convert after data has been entered in a particular field and if you convert from one format to another or one properties to another you will bring some errors in data and then might create lot of problems later on. So, it is always better to think first and then add the data inside your column.

(Refer Slide Time: 26:46)



Now, the last type of attribute is counts and amounts there sometimes that you want to show the total numbers and in at that particular field as a attribute and a count is actual number of features on the map and then, amount can be any measurable quantity associated with a feature. For example, number of student in a class and using a count or amount let us you see the actual value of each feature as well, as it is magnitude compared to other features. So, sometimes you want to store that data as counts or amounts and therefore, you go for this kind of attribute declaration.

(Refer Slide Time: 27:28)



Example, I am giving here that this is a district level map of Himachal Pradesh and this is census data of ninety one does not matter. Even if you are having two thousand eleven data scenario might be same and this the total counts in the each district of the people or the population district wise population. So, like here there are two districts and a there the range is also given and their total count is between this and then between this range similarly there are there are districts which are having say light green color and their total count is falling between these range.

So, here it becomes much easier to understand. If you use this kind of attribute that which type of you know which district having or which districts are having what kind of total number of populations. So, using a properly declaring your attributes and then bringing the data and later on you can use very well for different purposes. So, this brings to the end of this presentation.

Thank you very much.