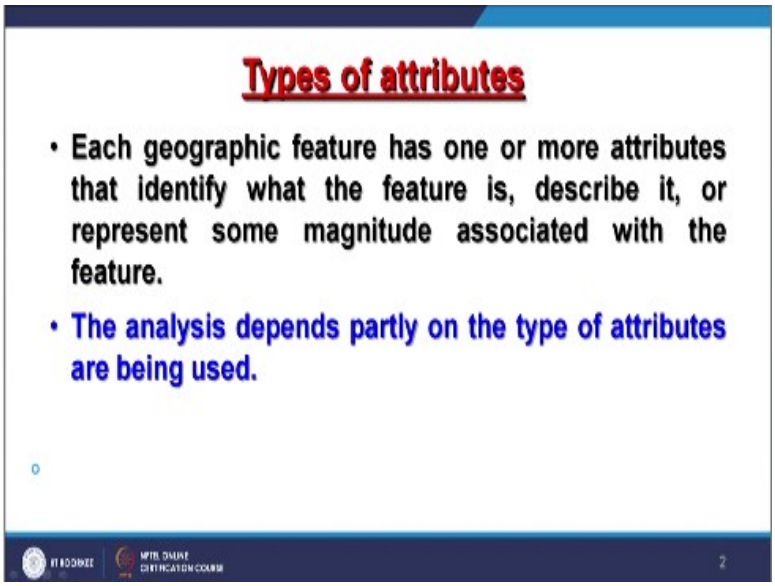
Geographic Information Systems
Prof. A. K. Saraf
Department of Earth Science
Indian Institute of Technology – Roorkee

Lecture – 08
Non-Spatial Data (attributes) and Their Types

Hello everyone! And welcome to new topic. In this, we are going to discuss non-spatial data, also called attributes data and their different types. If you recall that when we started discussion on data in GIS then we divided in 2 main categories. One is a spatial data and other one is non-spatial data. Spatial data: in which we had vector, raster and TIN. We have discussed in vector; all 3 vector entities point, poly-line/line and polygon.

And in raster, we have also discussed about grid and image, with 2 types of raster. And third, we have also discussed TIN. So in this one, we are going to discuss what is basically non-spatial data or attributes data; how it is handled, how it is stored and what are their types.

(Refer Slide Time: 01:15)

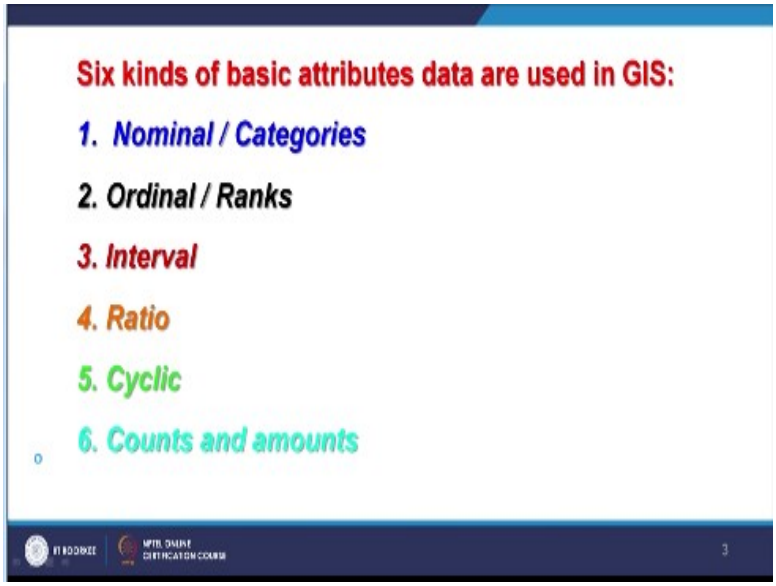


So, because as you know that each geographic feature can have one or more attributes. So, vector entities; point, line, polygon can have theoretically n-number of attributes but the raster can have only one type of attribute. TIN can have many because say triangle, it can have. This is how the data structures are. So, point data, line data and polygon data can have basically n-number of attributes.

So when we talk about attributes and types, mainly we are discussing basically about the vector attributes associated with vector files. Also, when I was showing demonstration and there also, I use the point data and perform some query on that. So because we had multiple attributes stored against each point and therefore, it was possible to do it. So that is what advantage of having as accurate and as rightly entered organize attributes in our GIS database.

Now as you know that a lot of analysis will depend on the type of attributes which are being used. Type maybe simple character strings, maybe numbers of different types- maybe integer number, maybe real number. If real number, what is the precision. Maybe some other types of data like date, time or you know like bearings or in degrees; in between 0 to 360 degree and so on so for.

(Refer Slide Time: 02:58)

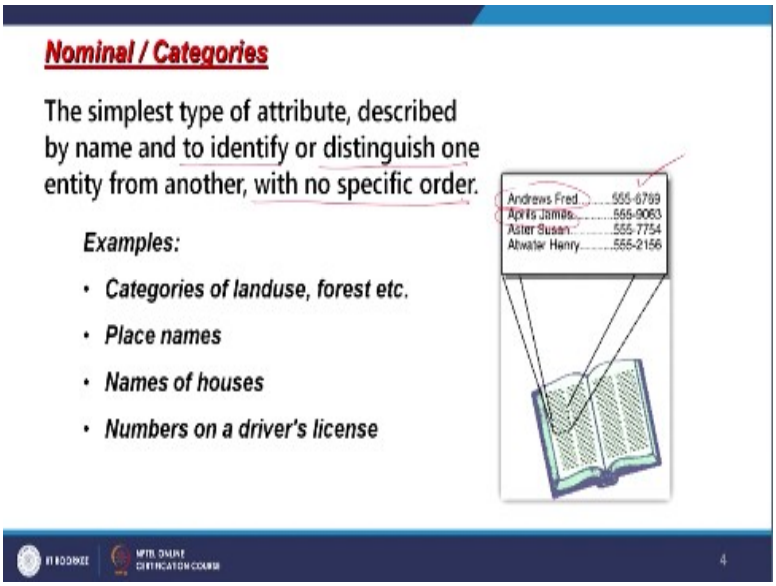


So, these are basically how the data will be stored or system will understand. So far in GIS, the 6 types of attributes have been implemented. The concepts of these attributes are available. First one is nominal or categories. Nominal is something like a proper noun in our English literature. And we will see also details against this. So let us first go one by one about the 6 kinds of basic attribute data. First one is nominal.

Sometimes in literature, you would find instead of nominal word, they use the categories also. Like in ArcGIS, the word category is used. Second is the ordinal or ranks data; there will be difference between nominal and ordinal, also called ordinal as ranks data. Also, we will see much today, then interval and then ratio then cyclic data.

Many times we have to use the cyclic data especially about bearing in other things and finally, counts and amounts. This counts and amount has been added quite recently. So in older literature, you may find only up to the five kinds of tools but now, you would find counts and amounts.

(Refer Slide Time: 04:19)



So one by one, we will go in detail. What is basically nominal data; it is the simplest type of attribute which describe name or identity or distinguish one entity from another with no specific order. Like here, there are names are given. Here, we may have say like telephone number but as definition says, in case of nominal or categories that just to identify or distinguish one from another and that without any specific order.

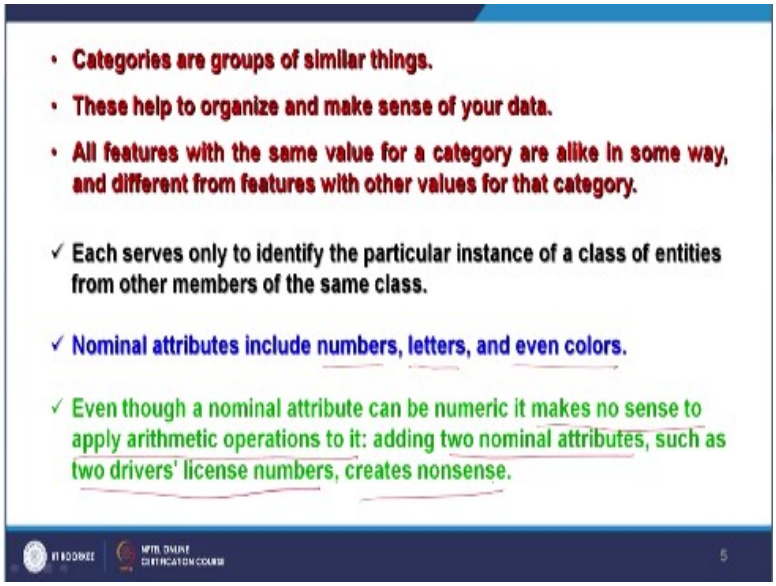
So with no specific order, it is just name. Though, one can argue that I can organize names in alphabetical order but that part is not inherent. You are doing an organizing alphabetical order. Otherwise, names in nature or such identities do not have any inherent order; that is the main point.

For example we can be categories of land use, like forest land or wasteland or water bodies or you know agricultural land or built up land. They are all categories of land use but there is no specific order between these categories. No one can say that you know, the wasteland is useless land and your water bodies are good. So, there is no order whatsoever. Same with the forest also; types of forests, open forests and the conifer forest and so on.

Name of places; again you can sort out in alphabetical order or any other order but inherently name of places does not have any specific order. Names of houses also, numbered on a driver's license. Now this is very interesting example. Though, it is a numeric value; driving license number is a number and that means I can perform arithmetic operations on this.

But if I perform on numbers, or like AADHAAR number then I may not have any results whatsoever because even husband and wives AADHAAR number may be completely different. So if I do it an average, it is not going to work.

(Refer Slide Time: 06:44)



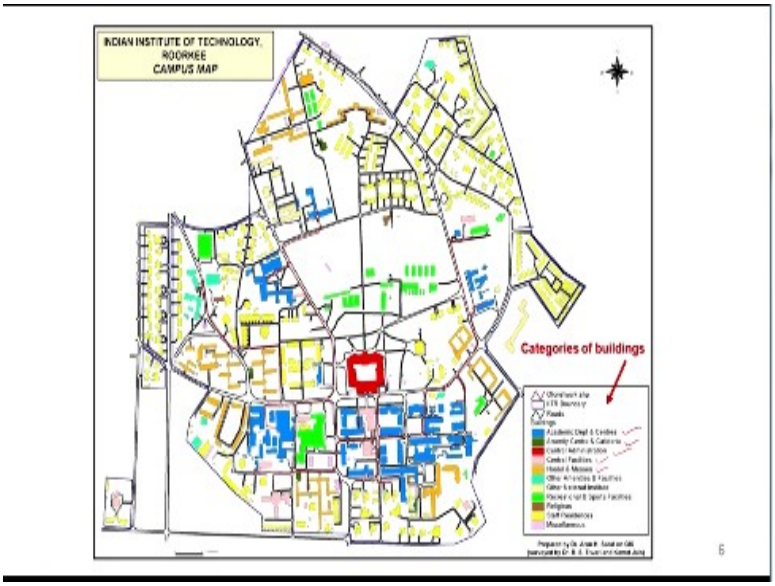
So, categories basically are groups of similar things; types of houses, type of forest, type of land, land use and so on so for. And these basically help us to organize and make sense in your data. When we give the name, it makes some sense. And it organize that each one is having their own name. And all features, here I am talking about vector entities with the same value for category is alike in the some way and different from features with other values for that category.

And each serves these categories, serves only to identify the particular instance of a class of entities from other members of the same class. So, 2 people are having 2 different names and therefore it is easy to identify who is who. And nominal attributes may include alphabets that any letters or even colors or numbers. Numbers can also be nominal attributes. But if we have declared in our database that particular field or column in our table or in excel is a nominal attribute then one should not try to perform any arithmetic operations.

Sorting is always possible. Sorting is not issue but performing arithmetic operations will yield basically no good results. For example if nominal attributes can be numeric as I have just mentioned, however it makes no sense to apply arithmetic operations to it. For example if I add 2 nominal attributes such as 2 driver's license numbers creates nonsense because they are cannot be an average license number or average AADHAAR number.

Though, the numbers are there and arithmetic operations theoretically are possible but practically there is no use of doing such exercises. So, all these things are come under this.

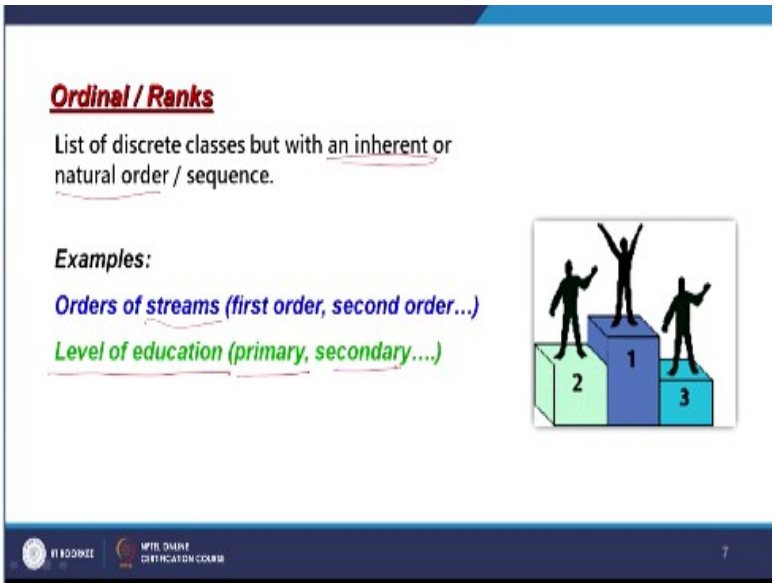
(Refer Slide Time: 08:45)



I am taking example of little older IITR map and these are the buildings which are being used for different purposes. And these have been categorized as like academic building, amenities, central administration, central facility, hostels and different colors and categories are given. Now these have been just ordered alphabetically but inherently all buildings are important and they do not have any inherent specific order.

And the map which I am showing later on, you would see the same map using different attributes from my database of this polygon information, I can reclassify. I can re-categorize my map accordingly. You would see the example very soon.

(Refer Slide Time: 09:42)

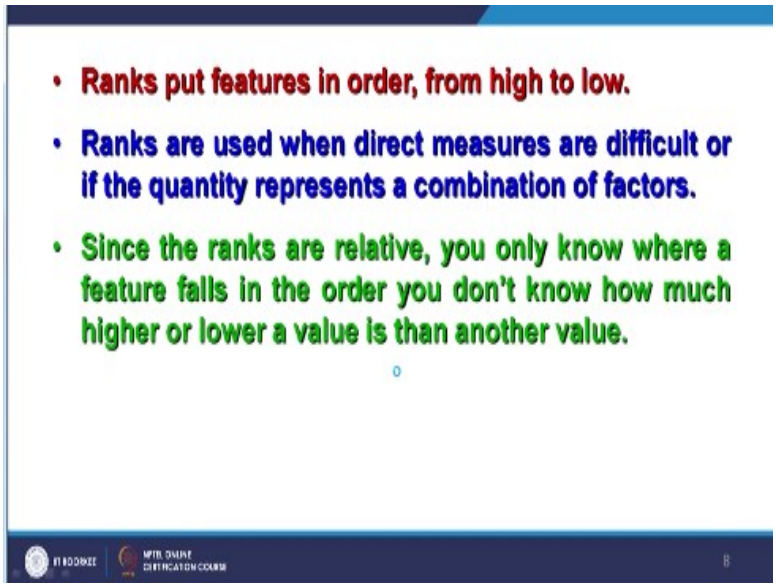


So, let us come to the next one or the second one which is ordinal or ranks. As name implies here that must be having some order. So it is again list of discrete classes but with an inherent order or maybe a natural sequence or order. For example like a stream network, we are having first order stream, second order stream and third order stream.

So if I say first order stream, it means it is having some meaning. It is having order because from there in that particular hill that river originate; originates from that point. And if I say third order then means first order and second order stream must have joined that is why it has become third. So, there is an inherent order or natural sequences there. Examples I have already just mentioned streams network, first order, second order and so on.

And level of education; like primary education, secondary education, undergraduate education, graduate education, postgraduate education and maybe doctoral or postdoctoral education. So, again there is an inherent order.

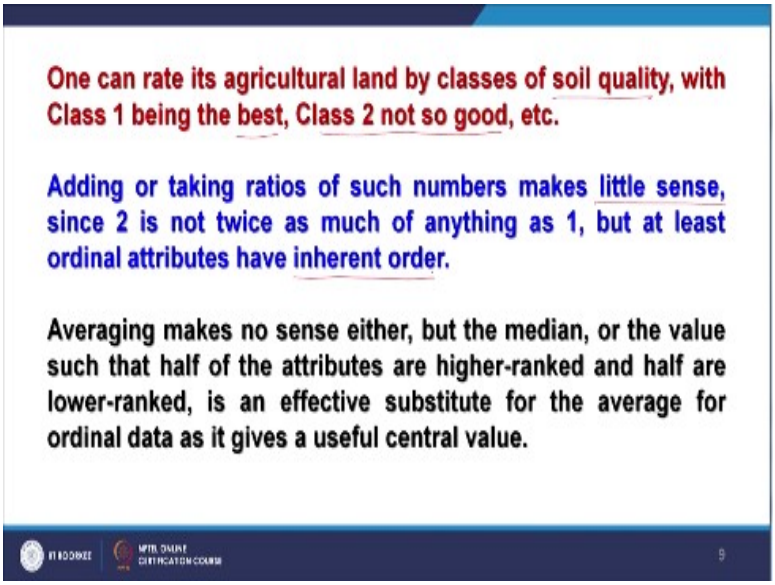
(Refer Slide Time: 10:58)



Now these you know, ordinal or ranks put features in order generally from high to low or ranks are used when direct measures are difficult and if the quantity represents a combination of factors then we use the ranks. And ranks are relative like a JEE examination, we also use the rank. So, these are relative. The first rank holder is having higher position than the second and the second one is having one lower position then first but having higher position then third.

So, it is a relative sense but as soon as I say ranks that means it has got inherent order. And if number is given then we can decide how much higher and how much lower.

(Refer Slide Time: 11:53)



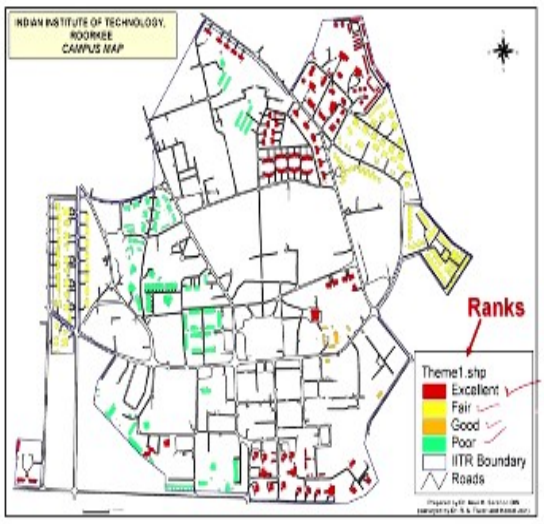
If number is not given then becomes sometimes difficult. So, one can rate his agriculture land for example by classes of soil quality or class being the best class, class 2 soil is not good

Etc. So, you can assign some order before you classify your data and adding or taking ratios of such numbers makes little sense but that means that quality 1 or class 1 divided by 2 or class 1 is adding by class 2 make no sense basically.

So in my attribute table, I may have for heading class or field name is class and I may have value 1, 2, 3, 4. But if I perform arithmetic operations, again it will not have any meaning. Though, it is having inherent order and again averaging makes no sense. But other maybe statistical analysis like median or value such that half of the attributes are higher rank or half are lower rank may have some meaning.

But generally if the data is ordinal or having ranks, arithmetic operations with such attribute data are not performed.

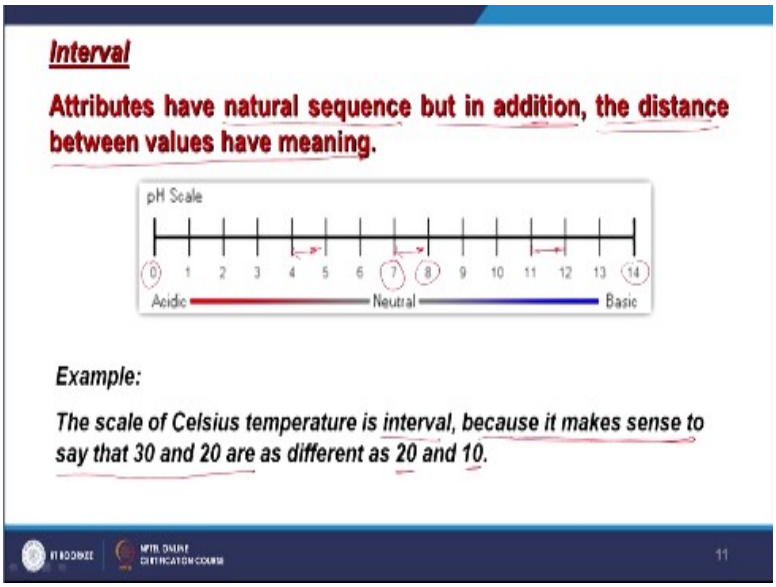
(Refer Slide Time: 13:21)



Now as I said earlier that the same map, I have ranked it and how I have ranked these buildings. That from construction or a condition point of view that is some of them is excellent, fair, good and poor. This just for you know analysis purposes; it does not represent the real picture currently. It is just for our understanding that how the same map using a different attribute which is having information that these individual buildings can have these 4 types of categories have been or ranks have been given.

And therefore, I can classify that map according to the ranks. And ranks are here depending on their condition whether they are in good condition, fair, excellent, poor or whatever.

(Refer Slide Time: 13:21)



Similarly the third attribute, now as we move higher and higher in attribute, the complication will keep adding like nominal does not have any kind of order. Whereas when we talked about ranks or ordinal, it is having inherent order. Now when we come about the interval, then interval has got the natural sequence. But in addition, the distance between the values have meaning.

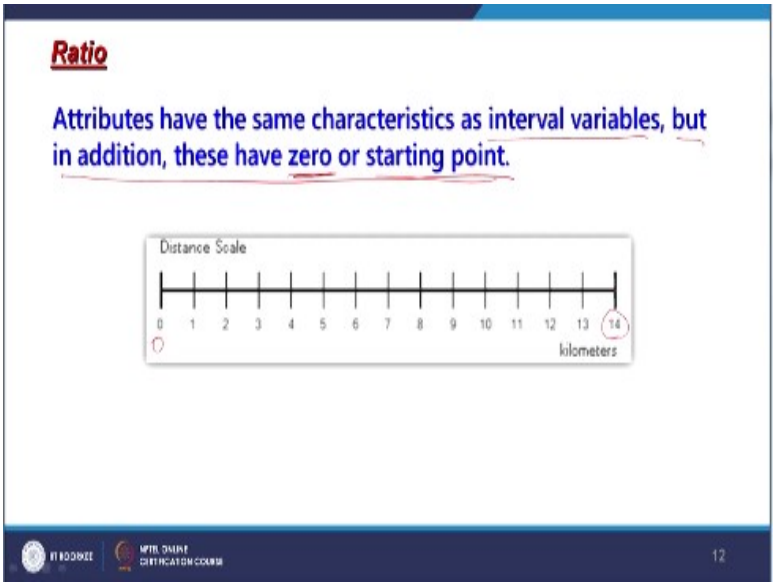
Like if I talk about ordinal then class 1 of soil in class 2 of soil, the distance between these 2 does not have any meaning but here 5 pH and 6 pH, 6 pH and 7 pH has got the meaning. So that is why in addition to the natural sequence, it is also having a distance and that has got the meaning. The example here is given about the pH. So as you know, pH 7 is considered as neutral and 0 is acidic and of course 14, we go for basic.

Now the meaning here that the distance has got meaning, so, 8-7=1 difference pH. The same, you can have 5-4=1 and same here also 12-11=1. So, the distance has got the same meaning in this particular example but has got the meaning and there is a natural sequence starting from 0 to 14; 7 neutral in case of pH. Similarly in case of like centigrade temperature that is a basically an interval attribute because it makes sense.

Now, some arithmetic operations are coming. So it makes sense to say that the temperature difference between 30, 20; 30 degrees centigrade and 20 degree are the same as temperature difference between 20 and 10. Human may feel a little differently but from measurement point of view, the distance is same and it has got the meaning. So if this would have been 20, 15 then only 5 degrees Celsius temperature difference.

So it is half of what between 20 and 10. That means some meaning is coming out and of course, natural sequence is definitely there.

(Refer Slide Time: 16:55)

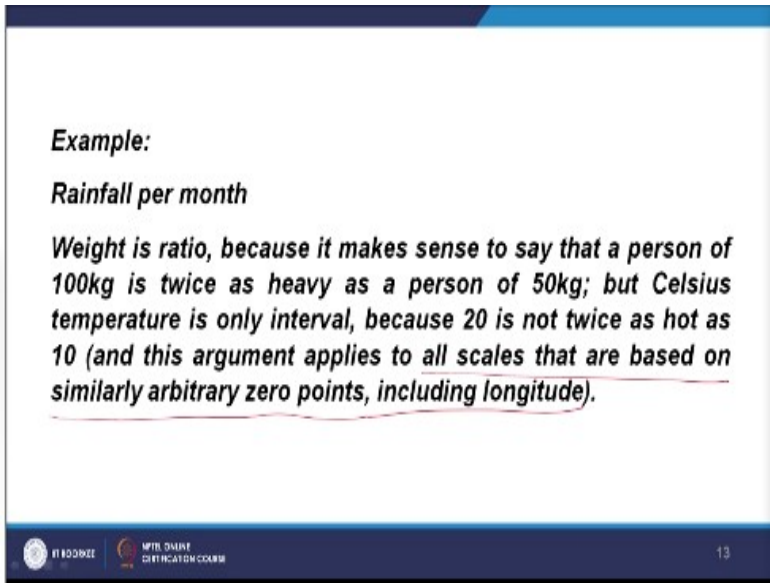


Now, the fourth type of attribute which is ratio and as you know that the ratios attributes have some characteristics. As interval variables little higher in hierarchy, it has got further some extra thing but in addition, these have 0 or starting point. Though the pH example we had 0 but there some other examples where we will not have 0.

So here the zero or the starting point is there, plus those conditions like inherent order and distance has got the meaning that is also there but in ratio that attributes had the same characteristics as interval variables. But in addition, these have 0 or a starting point. If I like in maps, we put a scale bar. This is that example. So, this is generally when a scale bar is put, we put the 0 and we put also 14 and scale is a ratio.

This one because if we say 50,000 then one unit on the map; maybe 1 centimeter equivalent to 50,000 centimeters on the ground. So it says a ratio and that is why it will have 0 values, starting point as well.

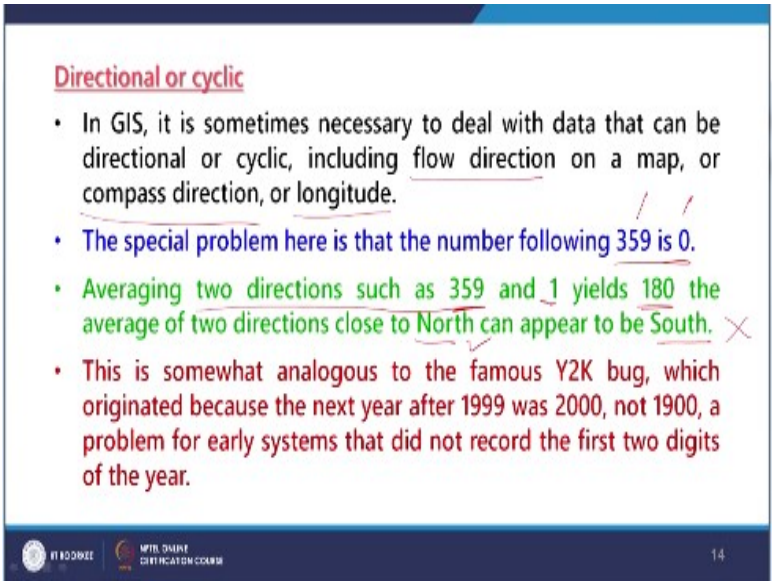
(Refer Slide Time: 18:22)



And examples like rainfall per month. So rainfall 0 is a starting point and that can be. Weight is also ratio. When we measure something, weigh something so, it is a ratio quantity and because it makes sense to say that a person of 100 kg is twice as heavy as a person of 50 kg. But in case of Celsius temperature, it is only interval because 20 is not twice as hot as 10 but you know this argument also applies to all scales that are based on similarly arbitrary 0 points including longitude as well.

This is important here longitude because we handle our data in geographic coordinates. So, there also, this is very-2 important.

(Refer Slide Time: 19:13)



Now fifth type is which is directional or cyclic. Lot many times especially in civil engineering domain or earth sciences where we have to record bearings like surveying; In

civil engineering, we record the bearing in which direction and maybe for some elevation in degrees also. And same like in earth sciences, we record dip, strike and plunge in everything. And they all are the directional data.

Whether they are in the horizontal plane that means say between north and full circle to the again north; 360 degree or maybe in a vertical plane starting from 0 to 90 degree. So, these are all directional. So in GIS, it is sometimes necessary to deal data that can be directional or cyclic. Here I want to mention that in earlier GIS systems, this capability was not there and we had a lot of difficulties when we started using GIS for directional data.

But now, most of you know popular GIS softwares are having capabilities of handling directional or cyclic attribute data very smoothly. Only thing while feeding or bringing data in your database, you have to declare the feed that it is going to be the direction that you know in a characteristics or format in terms of like excel, we call as a format of the cell.

That means the characteristics of the field that has to be declared that this data which I am going to bring is directional data. And this should be done before you bring the data in your database; attribute table because if you do it later on and declared then there may be some problems in your analysis or even in query. So, it is always better to create different fields having different types of attributes; having prior knowledge of what kind of data is going to be residing on different columns or fields and then you bring this.

So, flow direction is also directional data on a map or a compass direction. Wind direction is also cyclic data or longitude because longitudes which runs from north to south. And 0 degree longitude means that from London it overpasses, there is a GMT; Greenwich Mean Time also and then you go and take the entire circle of the earth and then you come back. That is why it is called cyclic. So, longitude is also cyclic.

The special problem is the numbers because after all these values are stored as numbers. 359 to next is zero because the next number is going to be the 0 or 360. But if I perform an arithmetic operation on these 2 numbers; 359 and 0 then I am going towards south that is incorrect. So, one has to be very-2 careful with directional data.

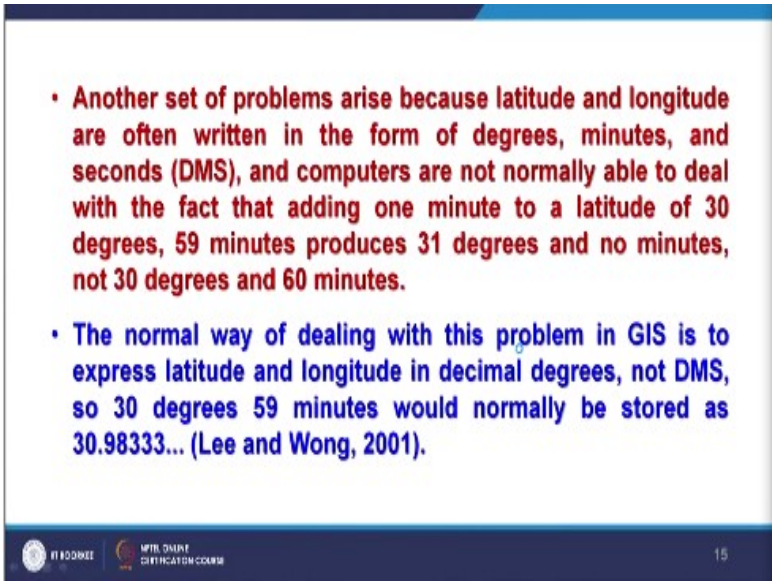
And that is why I mentioned that in the field itself, the characteristics of the fields should have that prior information that what kind of data is going to come in that column or in that database in that field. So, averaging 2 directions such as 359 and 1 yield 180 that means south. But basically, we know it is a directional data. So, 359 plus 1 equal to 0 that means 360 degree and it is the north, not the south.

So, one has to be very careful with the directional data. Nowadays, GIS are capable and no issue. This is a similar problem which was faced in the year 2000 when we changed the century from 1900 to 2000. At that time our computers systems, we are storing only 2 places for the year. That means instead of 1999, there we are storing only 99.

And if we go for 2000 that means that year would have been 00 and that can be 1900, that can be 2000, that can be 2100. So, this call the year 2000 or Y2K problem or bug and this were resolved by creating 2 more fields for the year in our computers. And quite easily, the problem was solved by the computer engineers that they redesigned their operating systems and could store information about the year in 4 digits rather than 2 digits.

So, one has to be very careful with the cyclic data because date or time is also cyclic data. After you know 23 hours and 59 minutes, again 0 hours comes, not 24 hours. So, that one has to be very-2 careful.

(Refer Slide Time: 24:47)



Another set of problems also arise because of this is in case of latitude, longitude which are written in forms of degree, minutes and seconds. That is why in GIS, we go for DD that is a

degree decimal and computers are not normally able to deal with the fact that adding 1 minute to a latitude of 30 degree 59 minutes produces 31 degrees and no minutes, not 30 degrees and 60 minutes because 60 minutes means 1 degree.

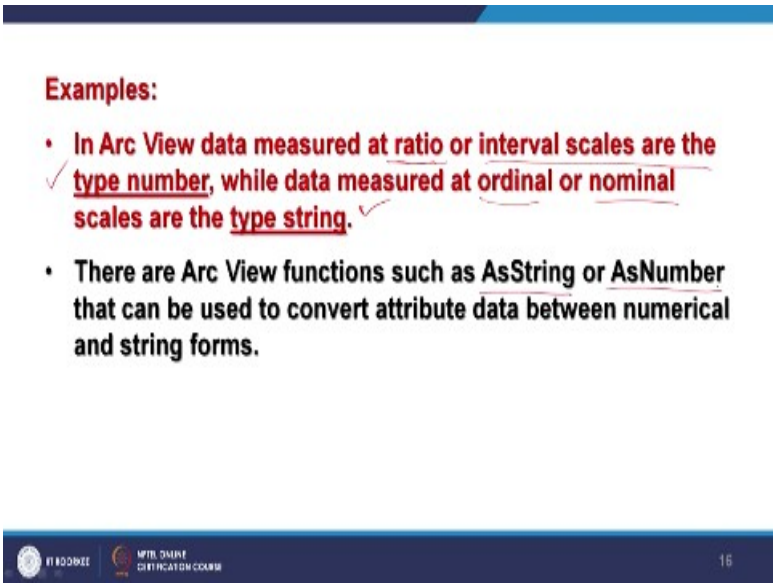
And it has to be transfer instead of 30; it should be 31 but if the data is being handled as like DMS degree minute seconds then problem will persist. So, one has to be very-2 careful and that is why I will repeat one more sort of rule which I framed in GIS. The first rule says that after each and every process steps taken in GIS, check for errors.

If you have fed the data into your database and you know this is a cyclic data, check it whether the system is correctly recording in that you know, rows or not against that field or column. And if you see that things are being recorded nicely then you go for the next otherwise correct it there itself because later on such attribute information cannot be corrected easily.

So, it is a big problem but there are solutions. Here what we store instead of DMS, we stored in you know degree decimal or in short, we also called DD. So, normal way of dealing with this problem is in GIS by cyclic data, was instead of not in a storing in DMS, we started storing in decimal degrees. For example is 30 degrees 59 minutes would normally be stored as 30 points.

It is basically a rescaling instead of 0 to 60, you rescale 0 to 100 and that can be done quite easily and that is why though that data may be stored for the spatial data and database, maybe in DD. This is how we store but anytime we want to display on the screen in DMS, we can do it very easily. And in the previous demonstration of GIS software, I showed these 2 things that you can see it in DD as well as in DMS. No issue.

(Refer Slide Time: 27:34)

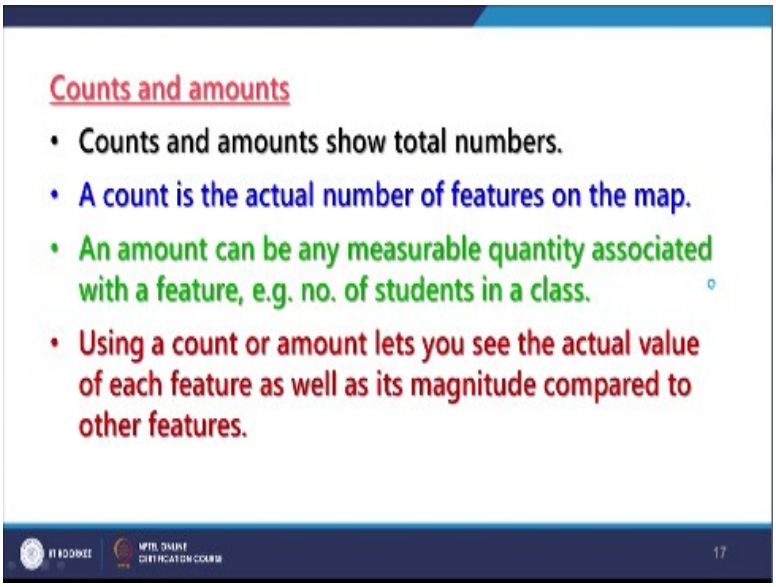


Now Arc View data, I have given the example. In Arc View, I also show to you that ratio or interval scales are type number; you know type of field which you would enter the data. So, when you are bringing ratio attributes or interval attributes, you would choose type number. And when you are having nominal or ordinal data then you are having type a string. These will matter when you go and perform the classification using any of the field's data which has been stored or attribute data.

So, all operations like query operations, like classification operations cannot be performed on all types of data. So, there will be some limit. Therefore that characteristic of each field has to be very carefully done. Now in Arc View functions as string also or as number that you have to declare and if you convert from one to another later on, once the data has been stored, there will be a lot of problems about decimal and other issues.

So, it is better always to decide which kind of data is going to come. Declare that field accordingly and then bring the data. Then hopefully things will be running very smoothly. Now last type of attributes or kinds of attributes which we have started handling in GIS in recent years is count and amounts. It is not that complicated.

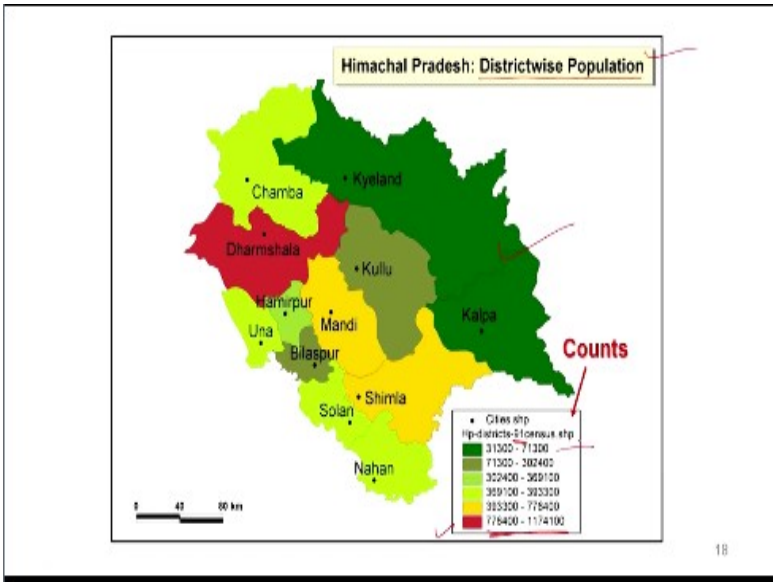
(Refer Slide Time: 29:12)



It is quite simple that counts and amounts show total number. So there is a separate attribute which will store the information about total numbers. And the count is actually a number of features on a map. Suppose there is a point map, I may have 1000 points. So, I would like to store in a separate field about this information. This helps while doing some statistical analysis over my attributes data.

So this field is very-2 helpful from that point. An amount can be any measurable quantity associated with a feature for example number of students in a class. And using a count or amounts; let us you see the actual value of each feature as the magnitude compared to other features. So if you are having data; depends on what kind of data you are going to handle and then counts and amounts can be very-2 important.

(Refer Slide Time: 30:13)



Now, I am taking another example from Himachal Pradesh and this is based on the census data and this is district by population as you can see, it is displayed here in the title of map. And total number of you know, people as per the 1991 census in each district; like a district which is having color, this is the range. So you know the total number and you know different districts as well.

Similarly here also, if I say this one then there are 2 districts or maybe 3 districts. If I remember correctly 2 districts; Kyeland and Kalpa, the 2 cities and districts are there so the range is like this. So you can have counts and amounts information as well. So, this brings to the end of our discussion about the kinds or types of attribute data. So we have completed now, different types of data which are used or handled in GIS.

We have discussed spatial types of data and we have also discussed non-spatial type data. We have discussed various types of spatial data and we have also discussed various types of attributes data or non-spatial data because as the good understanding we are having about the different types of data which is coming in our system and being organized because if you recall the definition of GIS, what is says that efficiently store the information.

If information is properly stored; efficiently store then only you can retrieve very efficiently and quickly. Sometimes if you start performing query and while developing that attribute database, you have not declared field as per their characteristics then in query, you will not get any results or may get erroneous results. May not be getting any results, may be still safe but getting wrong results may lead to you know, very disastrous results.

Therefore, it is always better to first understand what kind of data is going to decide in my database. Whether is spatial data or non-spatial data; which type of spatial data, which type of non-spatial data and accordingly. GIS database should be prepared. Thank you very much.