Spatial Statistics and Spatial Econometrics
Prof. Gaurav Arora
Department of Social Science and Humanities
Indraprastha Institute of Information Technology, Delhi

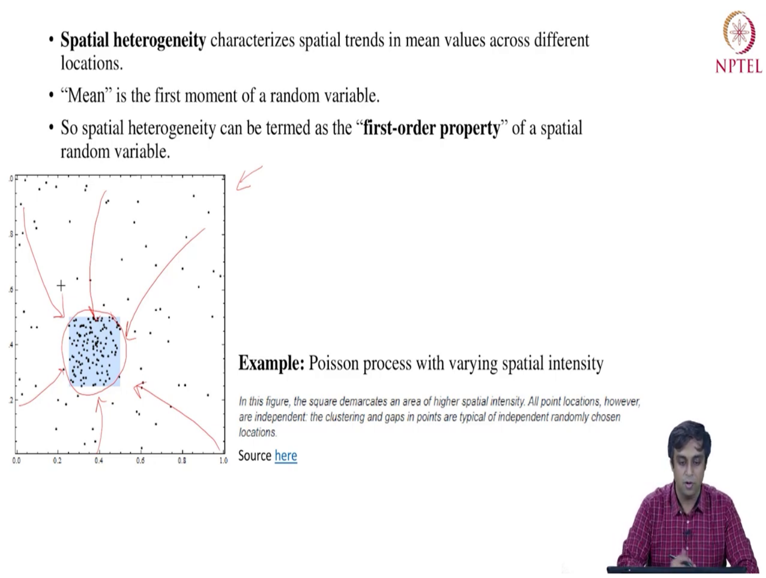
Lecture - 02
Spatial Patterns and Data Generating Process

Welcome everyone. We will now start the lecture 2 of Spatial Statistics and Spatial Econometrics. In this lecture I want to talk about characterizing spatial data. So, spatial statistics provides us two broad tools to delineate patterns in spatial data. The first is called spatial heterogeneity; the second is called spatial dependence.

The first that is spatial heterogeneity is more about spatial trends and we will look at it in more detail in a minute and spatial dependence as the name suggests is about dependence of how, you know, events are happening across space.

So, spatial heterogeneity characterizes spatial trends in mean values across different locations in space and when we talk about mean values; mean is the first moment of a random variable, right. So, we are viewing the world in as if you know the events that are happening are random events rights that is how statistics is entering the domain of understanding the happenings in the real world.

(Refer Slide Time: 01:30)

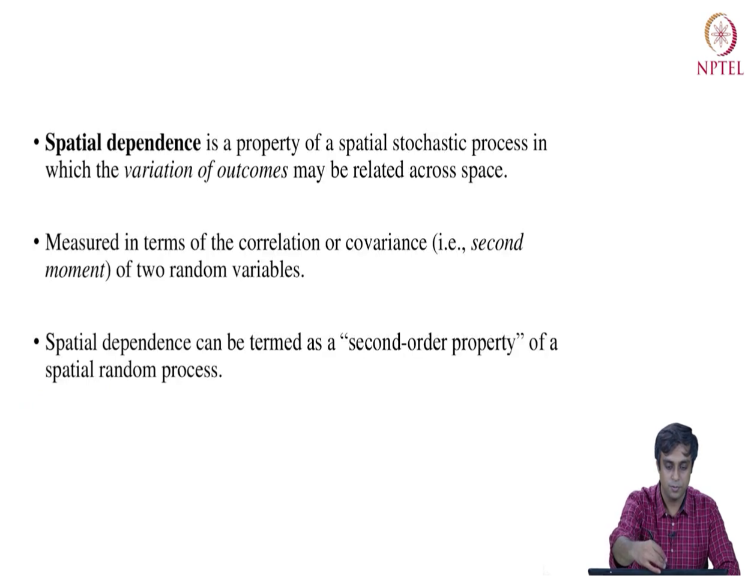


And then we talk about “spatial”, you know, what we are saying is how are these happenings across space and spatial heterogeneity is saying how different are the average you know intensity of events that are happening across space. So, mean being the first moment of random of a random variable, spatial heterogeneity can be termed as the first order property of spatial random variables. As an example what you see on the left bottom corner is a you know is a observation of signals at different intensities across space.

What you see is that there is this you know blue square box where the intensity of signals is quite high. So, the count of the observations or events happening is higher in the blue square than everywhere else in this box. So, what is happening is that if you think about a spatial trend of incidence of this events, we can draw, you know, we can draw you know a trend towards the blue square that is in the middle where the intensity of events becomes larger and larger as we move towards that box right.

So, there is some kind of a spatial trend that is occurring, you know, in this picture.

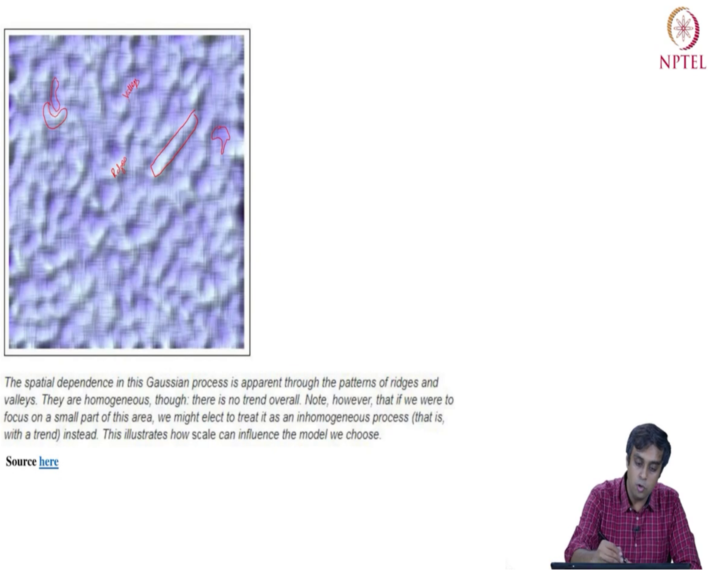
(Refer Slide Time: 02:54)



The second, you know, tool that we have is called spatial dependence. And spatial dependence is a property of spatial stochastic processes or spatial random processes in which the variation of outcomes may be related across space. So, now, how these outcomes vary across space may be different at different locations and we will look at a picture, so that we get an idea. It is measured in terms of correlation or covariance of two random variables that are, there, whose locations are identified in space.

So, now this is depending on the second moment of these random variables that is the covariances you know covariance metric. That is why spatial dependence is also sometimes termed as the second order property of a spatial random process.

(Refer Slide Time: 03:48)



To understand it a little bit better let us look at this picture in front of your screens. So, what you have is a combination of ridges and valleys. So, the whites is where you have ridges which are at a height and the blues where you have valleys which are at lower depth levels right that is how we are visualizing a combination of ridges and valleys from top, right.

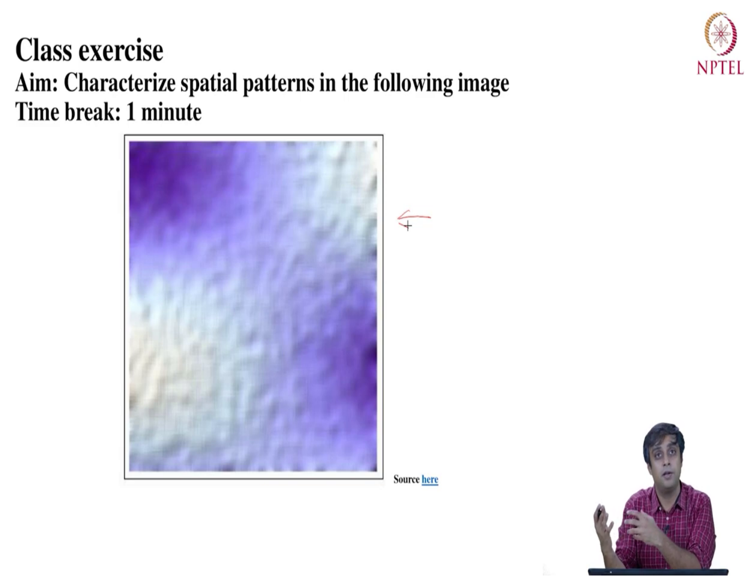
So, if you have a landscape where you have ridges in white right and valleys in blue right you have ridges and valleys occurring side by side. So, you have a ridge then you have a valley, then you have a ridge then you have a valley and so on and so forth.

The way ridges and valleys are distributed in space is quite random right there is no such pattern that ridges are occurring you know together in a cluster in a corner valleys are occurring you know there is a large valley or different combination of many valleys occurring in a different corner; no, we have ridges and valleys.

Ridges and valleys spread randomly across space. So, we do not really have any kind of spatial heterogeneity. So, to say in the trends of how ridges and valleys occurred they occur you know they are equally distributed across space, there is no specific trend. But what is happening is that the height of that height metric is clustered together at valleys at a certain level which is a lower level. They are dependent the height levels are dependent you know with each other for a given valley. So, you have lower height levels occurring together that forms a valley.

Similarly, higher height levels will occur together that is spatially in a spatially dependent manner to provide us a ridge. And that is how you have ridges where you have higher height levels occurring together or in a clustered way on space and lower height levels occurring together in a clustered way in space as valleys right. This is the fundamental understanding of spatial dependence right.

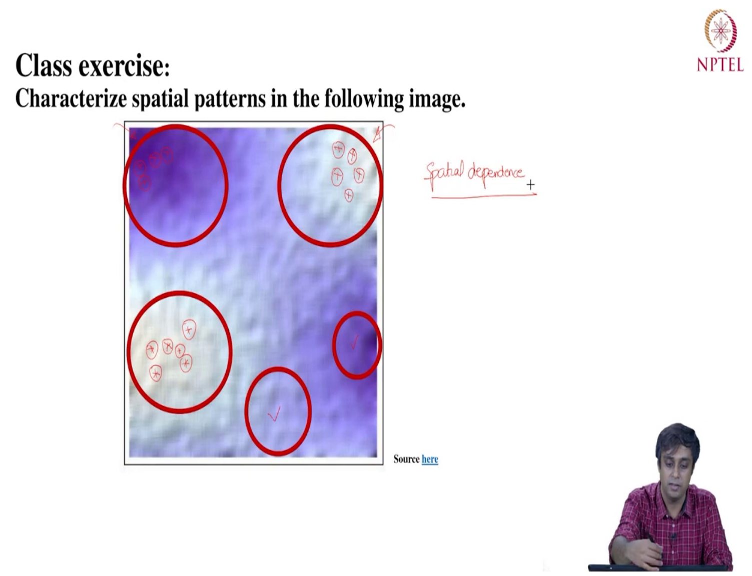
(Refer Slide Time: 05:58)



So, using these two concepts now, as a class exercise, I am going to give you 1 minute and then we will also solve this class exercise together, but I will give you 1 minute to now look at the picture on your screen, look at the image on your screen and figure out characterize the spatial patterns in it. Specifically, through the lens of spatial heterogeneity do you see any trends and spatial dependence ok. So, we will we only have two tools two instruments spatial dependence, second spatial heterogeneity.

Look at the picture and you know figure out how would you characterize the image in front of your screen we will give you we will come back in 1 minute and then solve the problem together. Welcome back. So, I hope you had a good look at the image in front of you and tried to characterize spatial patterns.

(Refer Slide Time: 07:01)



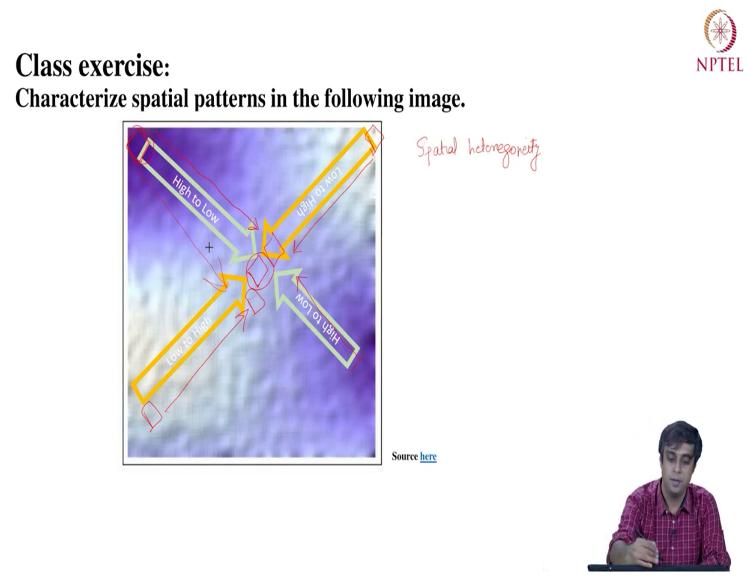
So, let us look at the spatial patterns that we can look, that we can find in this image. First of all focus on the red circles. What you see is that darker blue colors are occurring together as clusters on the you know north-west corner of the image. Similarly, lighter blue colors or whitish shade is occurring together and in clusters in the north east you know or the right top corner of this image.

This is a classic case of spatial dependence that is to say that the value that is, you know, noted here on this star on in the right top corner is quite similar to other values around it, right. They are all low blue less blue shade or low on the digital numbers shade right. In the north-west corner where you have darker colors again the value that is realized at any given point is quite similar to the values that are realized around it right.

So, there is dependence in what we observe in a neighborhood of any given stars in these red circles right. Similarly, you have this you know bottom left, you know, circle of white shades clustered together. So, the values in neighborhood are quite similar and you know we have two other circles where you can see similar patterns, right.

So, you can draw these regions in this image where the values exhibit quite a strong spatial dependence in the sense that whatever value we observe at a given point is quite similar to what is happening around in this space. So, this is the case of spatial dependence in the given image. So, the given image provides us an understanding or exhibits spatial dependence patterns that we can; that we can delineate.

(Refer Slide Time: 09:09)



The second thing that this, you know, image is able to also provide us an understanding about is spatial heterogeneity. As I said earlier spatial heterogeneity is about spatial trends. So, we should be able to see some kind of a trend of how the values are moving in space. So, quite clearly if we start at the north-west corner and we start coming to the center of mass of this image we are moving from high values to low values.

In other words we are moving from clusters of high values to clusters of low values. This is a trend in space. This is what we exhibit or we define as spatial heterogeneity. The average color shade in the region in north-west is darker than the average or the mean color shade of, you know, in this image at the center of mass right.

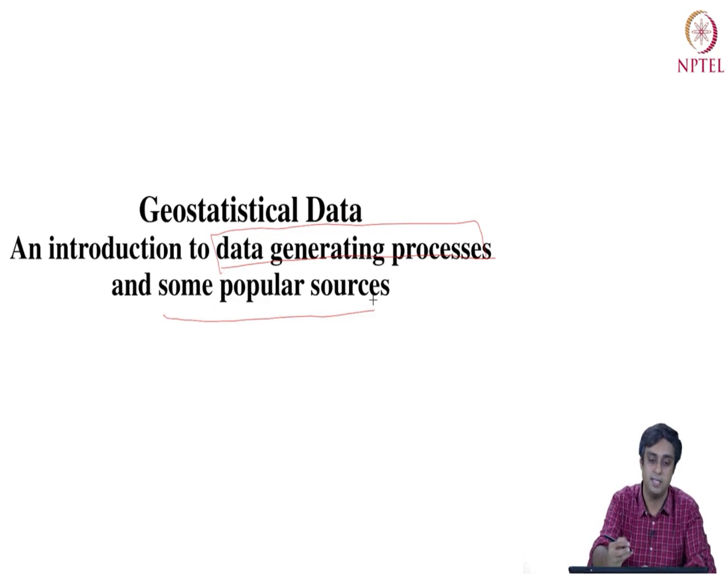
So, that is how there is variation in mean values in space and the way this is moving from high to low from north-west to the centroid from south-east. In fact, also to the centroid is exhibiting a one particular type of spatial trend. There is another spatial trend if we start from north-east and come to the centroid we are going from clusters of low or lighter shade to clusters of slightly darker shades right.

Similarly, if we go from southwest to the centroid we see this spatial trend of type 2 which goes which begins on one side at a lighter shade and goes up to a darker shade which might be exhibiting higher values or, you know, than the lighter shades right. So, now, we are able to characterize this image in two different ways, right. So, we started with this raw image here and we are able to, you know, delineate two different patterns here right.

One is of spatial dependence where we have clusters of similar values and different regions in this space that we can mark, identify and that gives us a better formal understanding of what is happening in this area. Second is we are able to delineate trends in how values are moving across space in a larger scale right. So, in general spatial dependence occurs at a smaller scale than spatial heterogeneity. Spatial heterogeneity asks us to traverse a larger region.

So, a larger spatial scale usually in general than what spatial dependence would ask us to do. Spatial dependence will be clusters so, things happening in localities. Spatial heterogeneity is what happening in two different regions in this image right. So, this is how you know we can characterize, you know, any given image using the tools of spatial statistics that is spatial heterogeneity and spatial dependence alright.

(Refer Slide Time: 12:24)

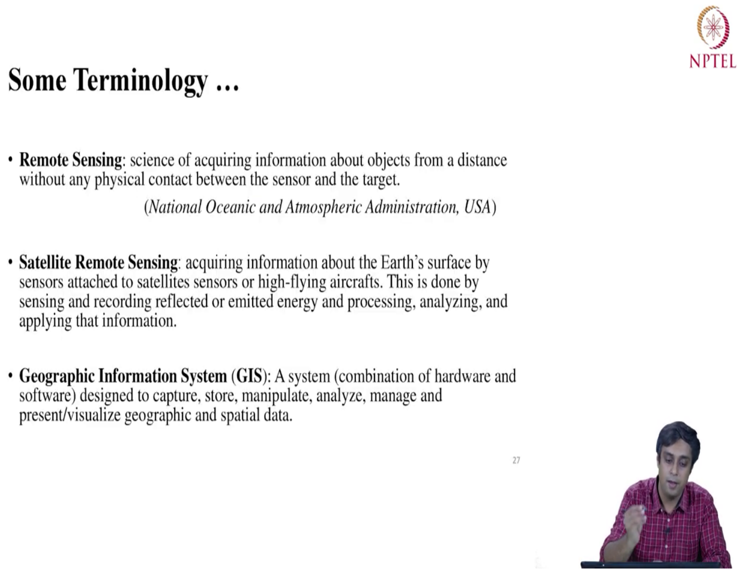


So, now we are going to sort of move forward in this course and do a detailed introduction of geostatistical data. What I want to talk about is, you know, provide an introduction before we move into the machinery of spatial statistics. We want to sort of talk about, where do these data come from, right. We have been looking at some of these very fancy maps you know they look very interesting they are stark they gives they provide us a lot of understanding of what is happening around us, but we want to understand where do they come from, right.

I mean unless we understand the data generating process you know we may not be able to fully appreciate the tool set or the toolkit that is spatial statistics right. So, we are going to talk about the DGP or the data generating process which is usually thought to be a very important component of statistical analysis. We should know where our data comes from. We should know how it is generated to be able to model it or to be able to make formal you know define formal statistics around it, right.

And we will also talk about we also look at some popular sources of data. So, that this the students you know who are taking this course they can go back and look at these courses and maybe just you know play with these data and try and get a understanding of what kind of, you know, what is possible with these data. Ok.

(Refer Slide Time: 13:51)



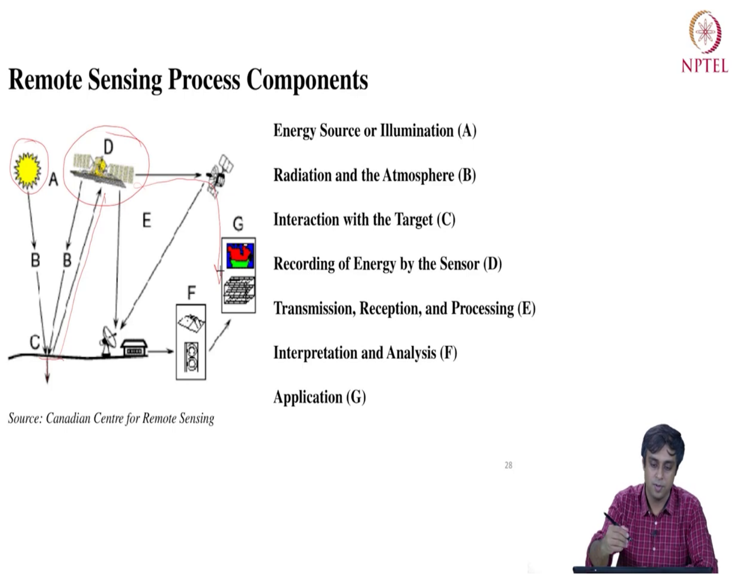
So, we will begin with some terminology; first important term terminology when it comes to geostatistical data is "remote sensing". Remote sensing is science of acquiring information about objects from a distance without any physical contact between the sensor; that is the one that is sensing information and the target that is the object of study.

So, the idea is that we are supposed to be studying an object without really touching it right, without even without really measuring it in physical sense we are supposed to do it from a distance. And what is satellite remote sensing well that is acquiring information about the earth's surface. So, now, the target or the object of study is the earth and we are using sensors, you know, attached to satellites. So, we are talking specifically about satellite sensors or high flying air aircrafts right.

So, this is done by sensing and recording reflected or emitted energy and processing analyzing and applying that information right. So, this is the second terminology, remote sensing, satellite remote sensing. The third term that I want to talk about here is GIS or the "Geographical Information System". GIS is a system which is a combination of both software and hardware designed to capture, store, manipulate, analyze, manage and present or visualize geographic and spatial data.

So, GIS is a software interface that allows us to visualize the data, store the data, analyze the data, manipulate the data right. So, GIS is really the physical entity that allows us to work with these data in practice right. The data by themselves are coming from sensors that are placed away from the object of interest or what we call as the target.

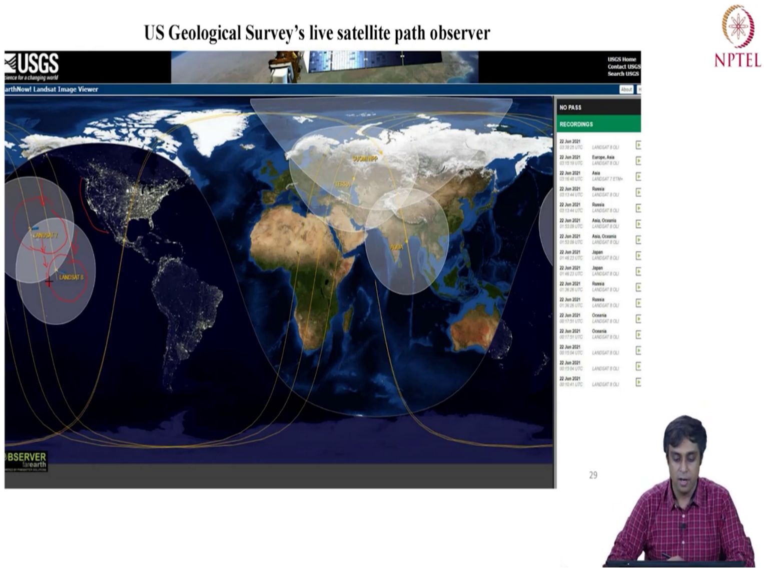
(Refer Slide Time: 15:40)



A bit of you know understanding of remote sensing process components you know what you see here is a source of light right. The source of light hits the target which is the earth's surface some of the energy observe is absorbed and some of it is going back to this satellite sensor which is roaming around in space right. So, the sensor collects this information in the in terms of energy received after reflection from the earth surface which is finally, received by a system on earth where its processed, analyzed, manipulated, stored and so on and so forth right.

So, the imagery that we saw towards the beginning of this course in the first lecture really went through all this process right. So, all this you know technology is going into getting that data to us, ok.

(Refer Slide Time: 16:34)

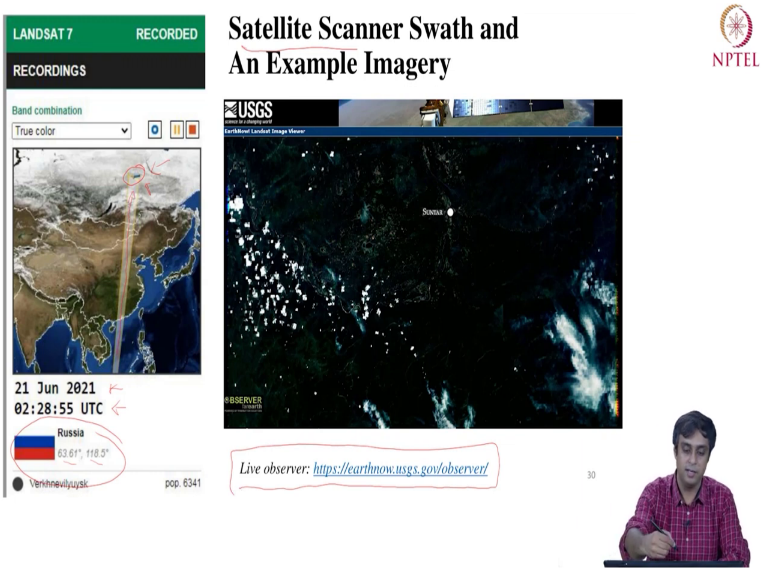


So, there are these live satellite path observers you know by the US Geological Survey that you can visit online through their Earth Now tool where you can actually see satellites you know moving in their paths in space. For example, in the image on your screen you know you can see LANDSAT 7 and LANDSAT 8 you know going around in the space towards in high seas in the west of the US California, you know, region right.

So, what is happening is that the satellite moves on its path and it as it moves on its path it clicks a picture right. So, it basically moves a step, clicks a picture, moves to the next step, clicks another picture, moves it to another step, clicks a picture. These pictures are then made available by the US Geological Survey or NASA or in India's case the Indian Space Research Organization or ISRO right.

India has a satellite called ResourceSAT right and its the sensors are specifically designed the Indian sensors are specifically designed to predict you know climatic events like cyclones right which are of the most important importance to the Indian region which experiences these events due to its large coastline right.

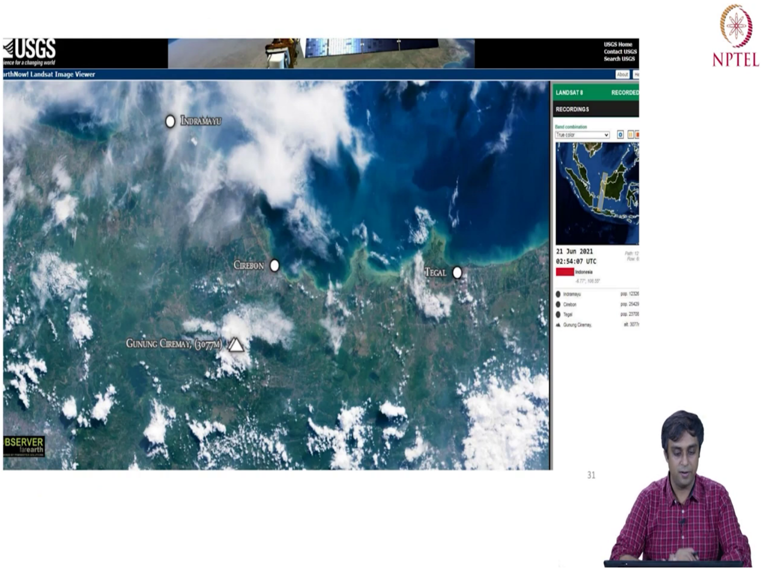
(Refer Slide Time: 18:05)



So, this is another picture, another example which gives you an idea of what we call as the "satellite scanner swath". So, you have a satellite which is going which is moving from you know south to north on the on Chinese on over China and then finally, it is over Russia you know at this point of time when it when this picture was when I when I took this screen grab this was at over Russia on 21st June 2021 given at a given time period with coordinate 63.61 and 118.5.

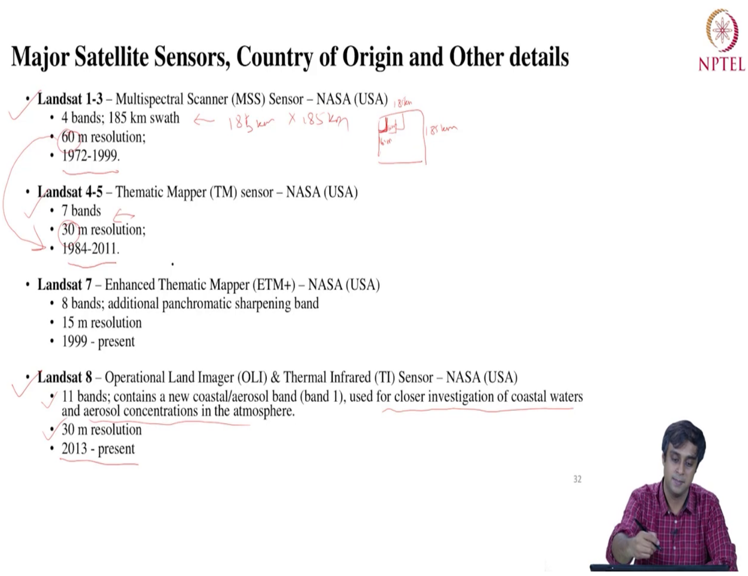
So, its on that spot it clicks a picture right it takes a picture the picture is of certain size for example, LANDSAT each picture will be of size 185 kilometer by 185 kilometers. So, each picture is so huge right, it covers that level of area on earth right. And you can actually visit this live observer on http://earthnow.usgs.gov/observer. Do visit this, it is pretty fun to go over this material.

(Refer Slide Time: 19:16)



Again another example you can take a look at it online when you visit the website. So, I want to just very quickly not spend too much time on this idea of major satellite sensors, the countries of origin are the other details. So, if we talked about sensors and satellite sensors.

(Refer Slide Time: 19:23)



So, I just wanted to give you some information about where these sensors are, what these and what are their different characteristics, when were they really released first right and you know and so on and so forth. So, first one I have here is LANDSAT. So, LANDSAT is

NASA's satellite sensor that is NASA is the space organization of the United States and NASA started its observation in 1972. So, 1972 to 1999 we have LANDSAT 1 to 3.

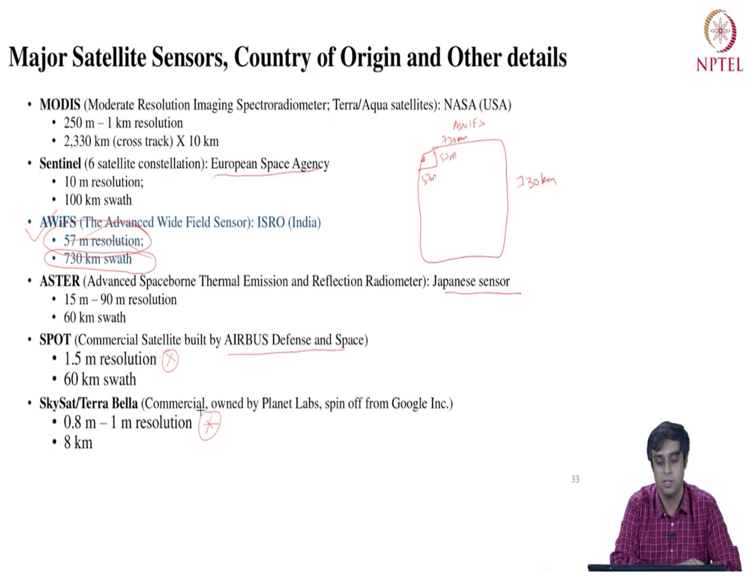
The sensor had 4 bands, four different bands, four different frequencies at which it was recording data and you know each swath was 185 kilometers; that means, each picture that it clicked was 185 kilometers by 185 kilometers. So, this is the size of each image that, you know, the sensor clicks at every instant that it is on space right. So, its continuously moving on space it comes to the same location every 16 days right.

LANDSAT 1 to 3 every 16 days it comes to the same location it clicks a picture the size of the picture is 185 kilometers by 185 kilometers. From 1984 to 2011 NASA had a more advanced sensor called LANDSAT 4 and 5 right. So, these had 7 bands that is why these were more this gave more finer details of you know the happenings on earth right.

The second difference between LANDSAT 1 to 3 and LANDSAT 4 and 5 is the resolution of the band. So, 60 meter resolution means that each pixel on this 185 kilometers 185 kilometers by 185 kilometers picture; each pixel that is the smallest entity at which information is available that is the resolution of the data is 60 meter by 60 meter this resolution increased from 60 meters to 30 meters when a new sensor was launched in 1984, right.

Sensors kept evolving, we are right now we right now have LANDSAT 8 going over us clicking pictures gathering information you know it went up in 2013 it still has a 30 meter resolution bit it has 11 bands. So, these are more sophisticated bands you know they are used for closer investigation of coastal waters, aerosol concentrations that is air pollution and idea of air pollution. So, you have more sophisticated technology as newer and newer sensors are sent to space.

(Refer Slide Time: 22:12)

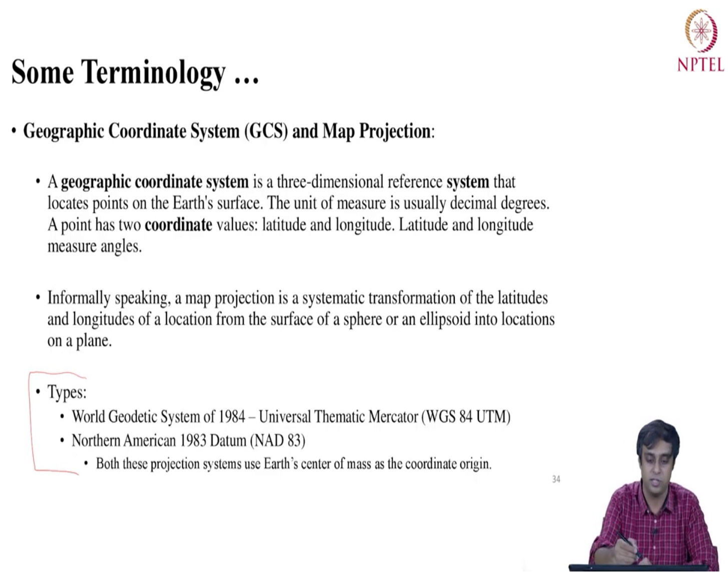


So, US is not the only of course, not the only country which sends out these sensors, you know. ISRO in India sent out the AWiFS sensor which is a; which is a lower resolution which is 57 meters resolution and a very large swath which is every image that AWiFS gets us is 730 kilometers by 730 kilometers ok. So, it is a huge image and every pixel is 57 meters by 57 meters. Why should we have a larger area coverage for an Indian satellite?

As I said the target the object of recording information is to predict climatic events like cyclones and these are large scale events. So, it is ok to have large you know areas covered at one point of time. In fact, it is more powerful for climate prediction purposes right. So, we also have things like Japanese sensors, we have sensors from European Space Agency there are also now private sensors right. So, you have SPOT which is AIRBUS Defense and Space. So, it is a commercial satellite what you see here is a very high resolution satellite.

So, 1.5 meter resolution, imagine the Indian satellite has the smallest pixel size as 57 meters whereas, the spot satellite has 1.5 meters it is an incredibly fine level of information is available through these private privately available you know sensors of course, they are paid they are not free freely available right. We also have satellite called SkySAT or Terra Bella which is a which is Google's sensors in space and of course, it is also you can see it is a very high resolution sensor ok.

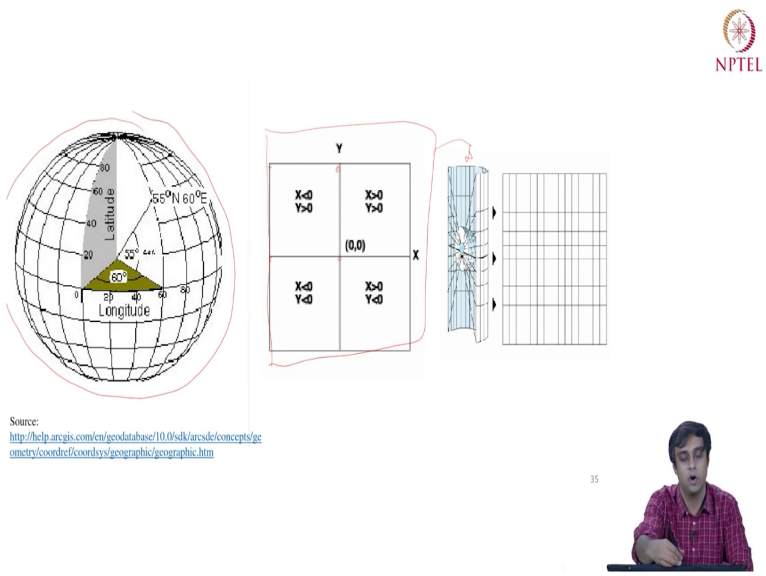
(Refer Slide Time: 24:00)



So, I want to very quickly sort of continue our data generating process. So, one more term that I want to introduce is called the "geographic coordinate system" and map projection which allows us to provide coordinates to the data. So, a geographic coordinate system is a three dimensional reference system that locates point points on earth's surface. The unit of measure is usually decimal degrees and a point has two coordinate values, a latitude and longitude.

Many of the students might be aware that when you have spatial data when you look at Google maps at every point that you point your finger or your you know or your you know a marker, you can you actually see coordinates. So, these coordinates are nothing, but a map projection which is a systematic transformation of latitudes and longitudes of a location from the surface of a sphere or ellipsoid into locations on a plane. And there are different types of systems that are used to create these projections.

(Refer Slide Time: 25:01)

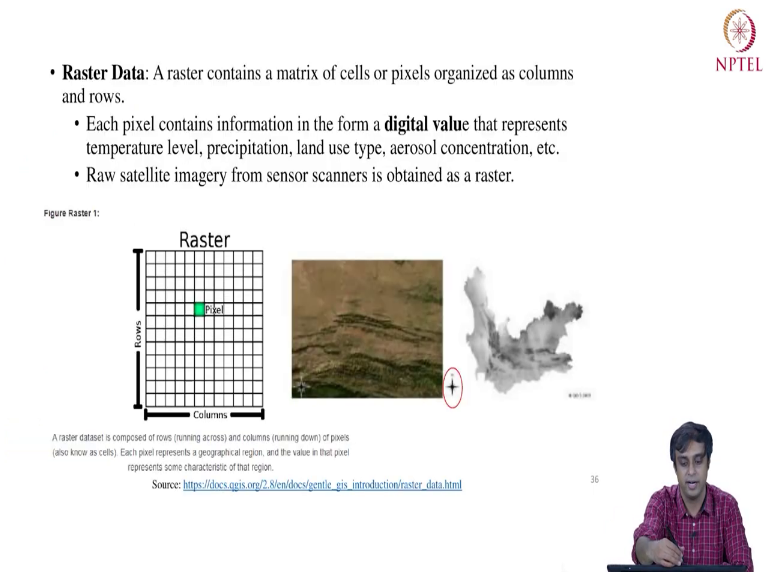


A picture will help give you an understanding of what is a geographic you know projection system. So, you have a; you have a sphere on which you want to; you want to provide these coordinates. What you do is you take a torch on this side or a light source on this side and project how the sphere would look like on a two dimensional piece of paper.

Once you do that you know you can then assign x y coordinates like we know how to do that on a two dimensional space and then finally, roll it back to the sphere to assign the coordinates on that sphere right.

So, what you are doing is you are bringing this 3D information from a light source to a 2D and on 2D we are well aware of how to assign coordinates on a graph which is x y coordinates right. And we call them latitudes and longitudes and we assign them back to our sphere, roll them back to our sphere and that is how we have geographic coordinates.

(Refer Slide Time: 26:01)



Alright. So, that is it for this lecture. In the next lecture we will study spatial data structures, specifically vector data and raster data and we will see how spatial data are stored in different formats and I look forward to having you there.

Thank you for your attention.