

# Lecture 3: Data Collection in Geography



# Announcements

1. This Thursday (Jan 24<sup>th</sup>), we will be doing some data collection outside in groups. Please bring something to write on (e.g. a notebook or a tablet) and dress for the weather.

## Reminders:

1. Homework #0 is due this Thursday (Jan 24<sup>th</sup>)
2. The NGA is on campus recruiting tomorrow
3. Class is canceled next Tuesday, Jan 29<sup>th</sup>



# Definition of Data

- Information that can be analyzed and conclusions can be drawn from
- Collection of data presented in tabular form
  - **Observations:** individual measurement in a set (varies over what)
  - **Variable:** property or characteristic that is being measured (what varies)
  - **Data value:** attribute measurement (how it varies)
  - **Scale:** time and space was the attribute measured

Year	Floods	Cyclones/floods	Hailstorm	Earthquakes	Total
1990-1991	1,320	979	—	—	2,299
1991-1992	1,185	304	—	768	2,257
1992-1993	1,193	497	—	—	1,690
1993-1994	1,690	318	—	7,938	9,946
1994-1995	2,038	247	59	—	2,344
1995-1996	2,072	361	31	—	2,464
1996-1997	2,069	1,719	40	—	3,828
1997-1998	1,560	216	247	39	2,062
1998-1999	2,567	1,292	—	106	3,965
Total	15,694	5,933	377	8,851	30,855

Source: Ministry of Agriculture, Government of India

Table I.  
Human lives lost due to  
various disasters in India

Year	Floods	Cyclones/floods	Hailstorm	Earthquakes	Total
1990-1991	1,320	979	—	—	2,299
1991-1992	1,185	304	—	768	2,257
1992-1993	1,193	497	—	—	1,690
1993-1994	1,690	318	—	7,938	9,946
1994-1995	2,038	247	59	—	2,344
1995-1996	2,072	361	31	—	2,464
1996-1997	2,069	1,719	40	—	3,828
1997-1998	1,560	216	247	39	2,062
1998-1999	2,567	1,292	—	106	3,965
Total	15,694	5,933	377	8,851	30,855

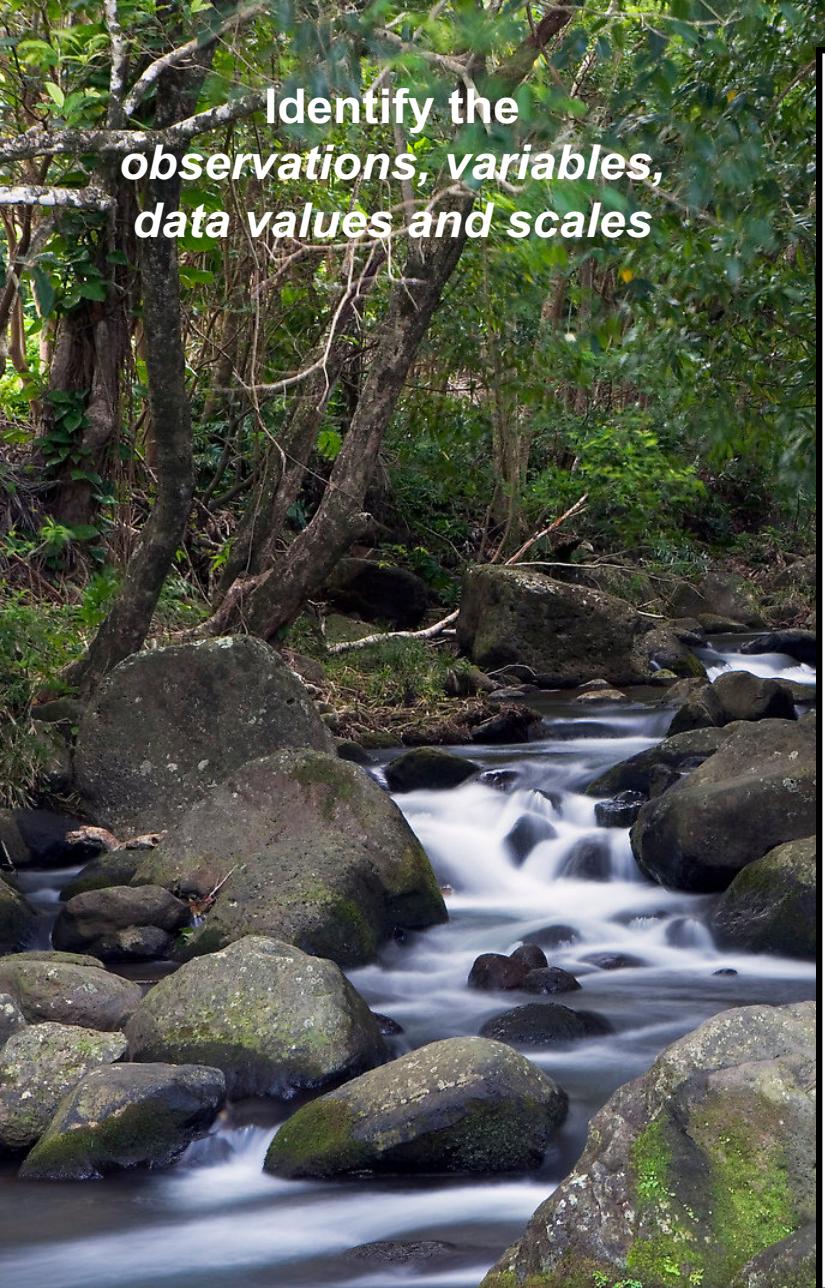
**Source:** Ministry of Agriculture, Government of India

**Table I.**

Human lives lost due to various disasters in India

**Identify the *observations, variables, data values and scale* in the above table**

# Identify the observations, variables, data values and scales



**Table 3.2—Physical variables in relation to stream order in 2 Oregon watersheds<sup>a</sup>**

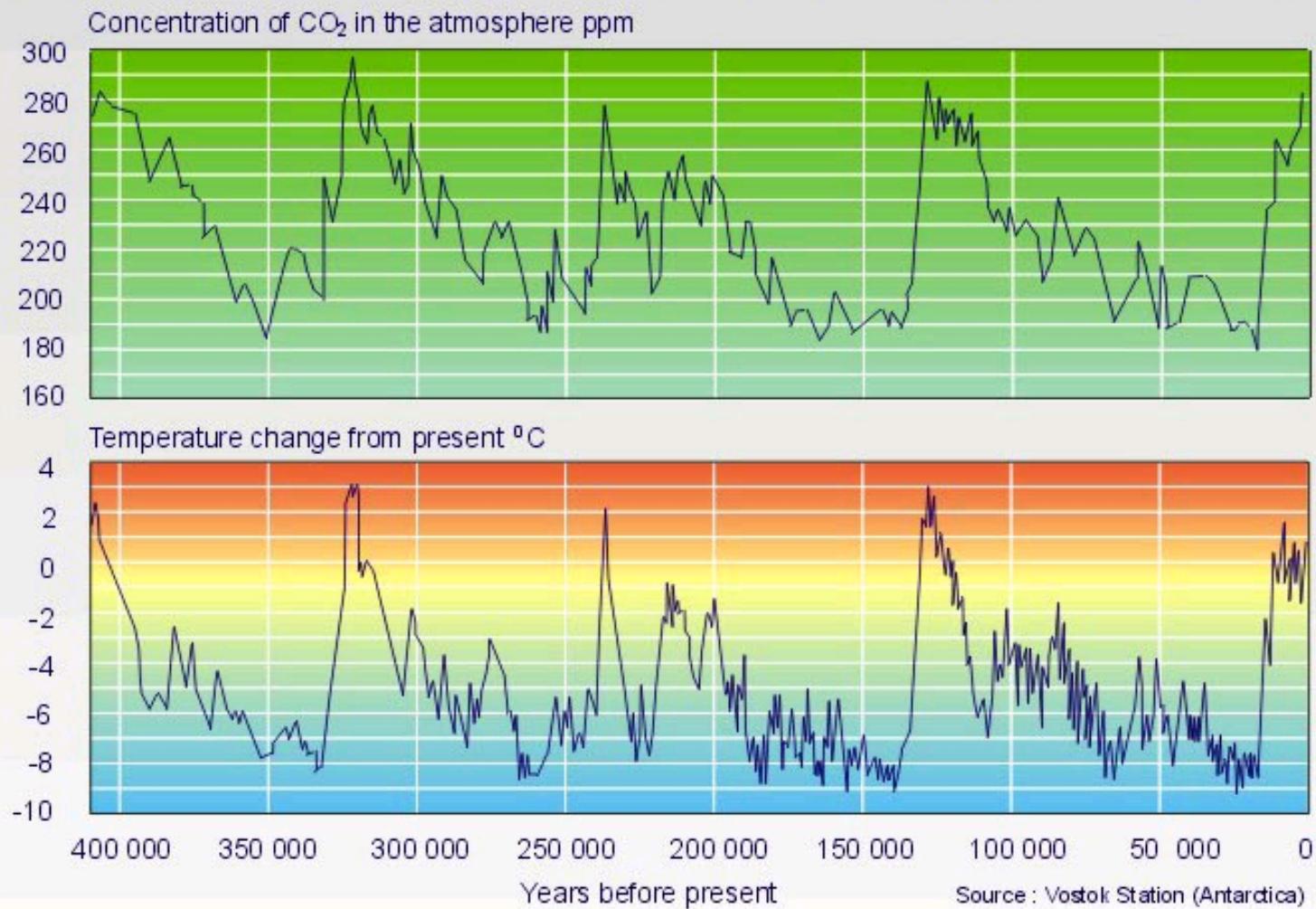
Watershed type and physical variable	Unit	Stream order			
		2	3	4	5
<b>Coastal watersheds:</b>					
Gradient	Percent	18	6	4	3
Width	Feet	2	5	10	11
Depth	Inches	8	16	16	33
Pool	Percent	22	27	31	37
Riffle	Percent	78	73	69	63
Spawning gravel	Square yards per mile	121	455	1,176	63
<b>Cascade watersheds:</b>					
Gradient	Percent	11	9	6	3
Width	Feet	4	7	9	24
Depth	Inches	2	4	8	12
Pool	Percent	55	56	58	49
Riffle	Percent	45	44	42	51
Spawning gravel	Square yards per mile	73	125	376	1,321
<b>Bottom substrate—</b>					
Bedrock	Percent	—	—	—	—
Boulder	Percent	—	—	—	—
Cobble	Percent	—	—	—	—
Rubble	Percent	—	—	—	—
Gravel	Percent	13	16	18	20
Fines	Percent	3	4	5	3

— = not available.

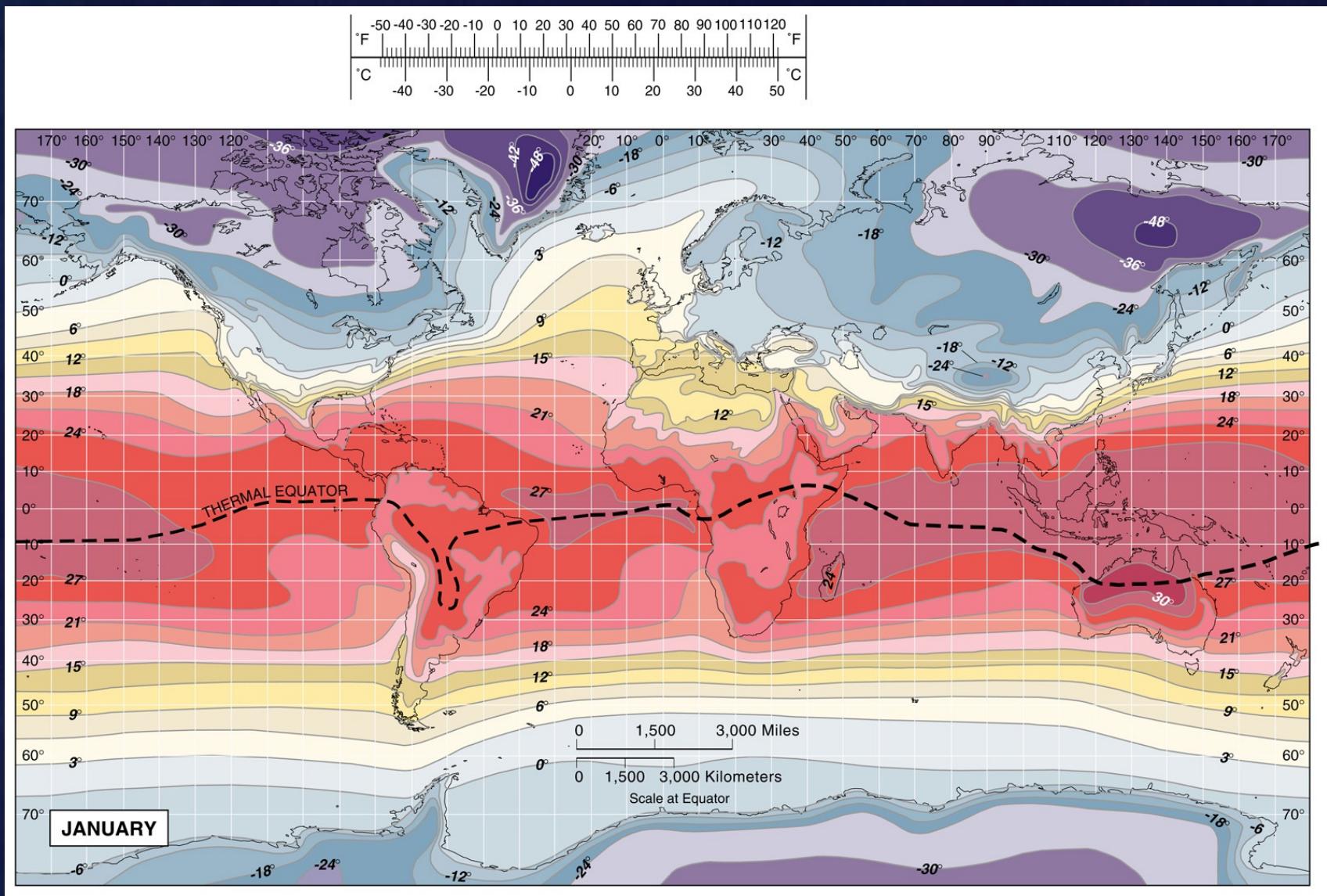
<sup>a</sup> Adapted from Boehne and House (1983).

## Identify the *observations, variables, data values and scales*

### Temperature and CO<sub>2</sub> in the atmosphere over the past 400 000 years



## Identify the observations, variables, data values and scales



### ***Primary Data:***

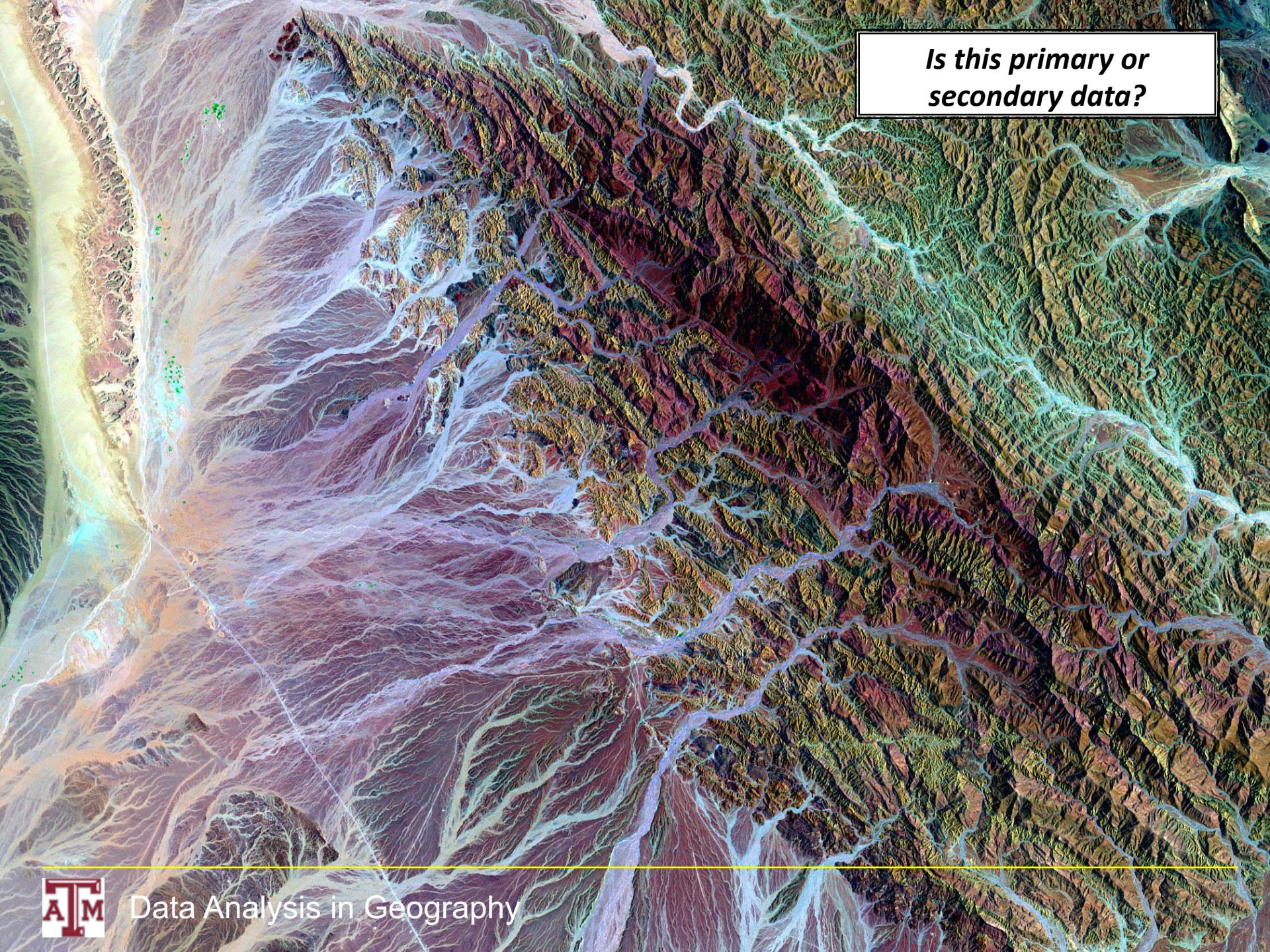
- Directly sampled/measured
- Specifically tailored to the research question being explored
- Requires considerable time and effort to collect
- Raw



### ***Secondary Data:***

- Sampled and measured by someone else
- Not necessarily collected for the research question
- Readily available but in archives
- Processed and organized





*Is this primary or  
secondary data?*



# Metadata

- Data can stand on its own without description or context, but eventually the data may be used by other people
- **Metadata:** provides information about data, including the processes used to create, clean, check and produce it
- Important source of information for explaining variation within and between samples

**Table 3.2—Physical variables in relation to stream order in 2 Oregon watersheds<sup>a</sup>**

Watershed type and physical variable	Unit	Stream order			
		2	3	4	5
<b>Coastal watersheds:</b>					
Gradient	Percent	18	6	4	3
Width	Feet	2	5	10	11
Depth	Inches	8	16	16	33
Pool	Percent	22	27	31	37
Riffle	Percent	78	73	69	63
Spawning gravel	Square yards per mile	121	455	1,176	63
Bottom substrate—					
Bedrock	Percent	2	16	14	9
Boulder	Percent	11	15	15	14
Cobble	Percent	26	15	17	19
Rubble	Percent	9	10	15	11
Gravel	Percent	36	29	27	27
Fines	Percent	16	16	12	20
<b>Cascade watersheds:</b>					
Gradient	Percent	11	9	6	3
Width	Feet	4	7	9	24
Depth	Inches	2	4	8	12
Pool	Percent	55	56	58	49
Riffle	Percent	45	44	42	51
Spawning gravel	Square yards per mile	73	125	376	1,321
Bottom substrate—					
Bedrock	Percent	—	—	—	—
Boulder	Percent	—	—	—	—
Cobble	Percent	—	—	—	—
Rubble	Percent	—	—	—	—
Gravel	Percent	13	16	18	20
Fines	Percent	3	4	5	3

— = not available.

<sup>a</sup> Adapted from Boehne and House (1983).

# Metadata

- Facilitates the understanding and use of data
- Data definitions: components, scope and purpose of the data
- Method of sampling: how was sampling completed and procedures were used
- Data quality: accuracy, precision and validity of the data
- Data dissemination: how the data can be distributed
- Geographic data: spatial information

Table 3.2—Physical variables in relation to stream order in 2 Oregon watersheds<sup>a</sup>

Watershed type and physical variable	Unit	Stream order			
		2	3	4	5
Coastal watersheds:					
Gradient	Percent	18	6	4	3
Width	Feet	2	5	10	11
Depth	Inches	8	16	16	33
Pool	Percent	22	27	31	37
Riffle	Percent	78	73	69	63
Spawning gravel	Square yards per mile	121	455	1,176	63
Bottom substrate—					
Bedrock	Percent	2	16	14	9
Boulder	Percent	11	15	15	14
Cobble	Percent	26	15	17	19
Rubble	Percent	9	10	15	11
Gravel	Percent	36	29	27	27
Fines	Percent	16	16	12	20
Cascade watersheds:					
Gradient	Percent	11	9	6	3
Width	Feet	4	7	9	24
Depth	Inches	2	4	8	12
Pool	Percent	55	56	58	49
Riffle	Percent	45	44	42	51
Spawning gravel	Square yards per mile	73	125	376	1,321
Bottom substrate—					
Bedrock	Percent	—	—	—	—
Boulder	Percent	—	—	—	—
Cobble	Percent	—	—	—	—
Rubble	Percent	—	—	—	—
Gravel	Percent	13	16	18	20
Fines	Percent	3	4	5	3

— = not available.

<sup>a</sup> Adapted from Boehne and House (1983).

# Common Types of Data Collection in Geography

- **Physical measurement:** direct measurement of physical attributes of the natural and social environment
- **Behavioral observations:** direct observations of the actions and activities of individuals
- **Archival:** use of existing records that others have collected (primarily) for non-research purposes
- **Explicit measurements:** individual answers to a survey that expresses an option or belief about something
- **Modeling:** physical or numerical modeling of process and response in natural and social systems



A surveyor in a white t-shirt and yellow safety vest is operating a total station at a construction site. In the background, a city skyline with several skyscrapers is visible under a cloudy sky. A yellow excavator and a blue pickup truck are also present in the scene.

# Physical Measurement

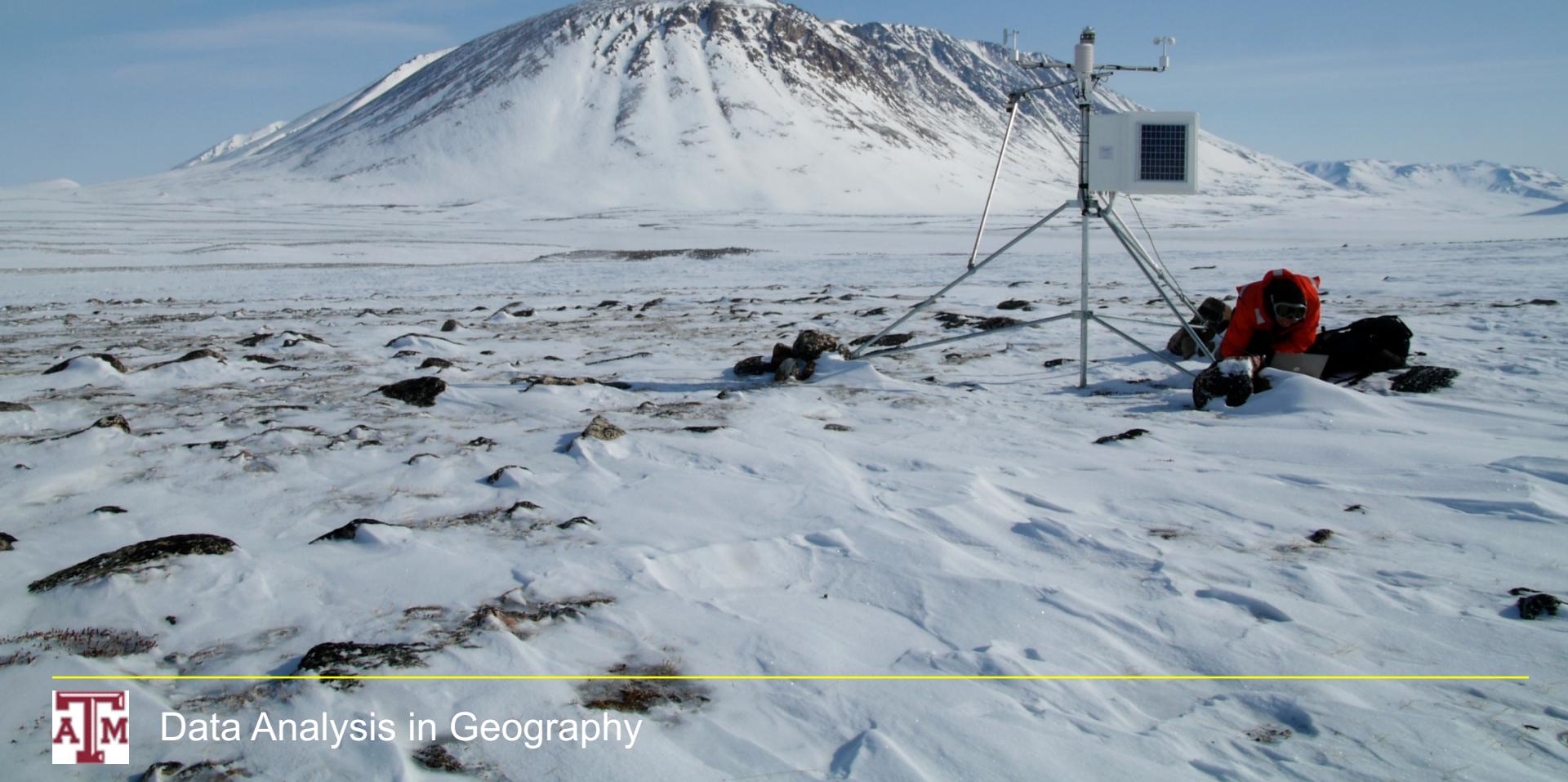
# Physical Measurement



# Physical Measurement



# Physical Measurement



# Physical Measurement



# Behavioral Observation



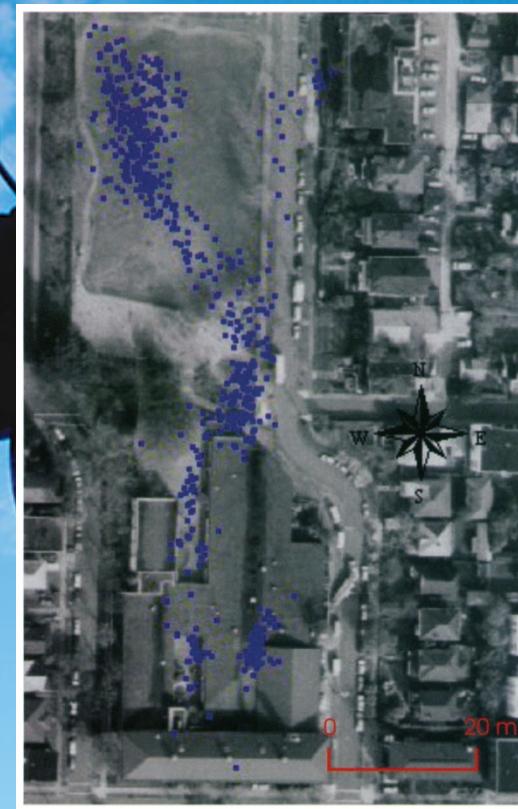
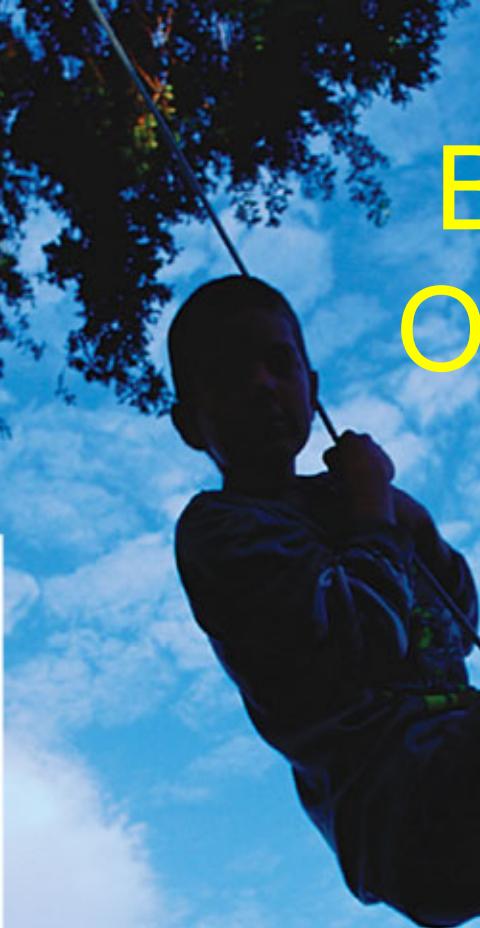
# Behavioral Observation



# Behavioral Observation



# Behavioral Observation



# Archival Measurement



# Explicit Measurement



# Laboratory Studies



*Specially designed setting that allows for **physical control** of the environment*

*Does this artificial setting simulate a natural river*

# Controlled Field Studies



**Experimental:** factors under consideration are controlled in order to isolate their effects on the variables of interest

# Uncontrolled Field



***Non-Experimental:*** no control is exercised over the factors that may affect the variables of interest but done with experimental rigor

# Qualitative Data

- Qualitative: variables are placed into discrete categories that are non-numerical
- Nominal data: categorical data that can be counted but not ordered or measured
- Ordinal: ranked data that can be counted and ordered but not measured
- Limited and different types of statistical tests that can be completed on nominal and ordinal data

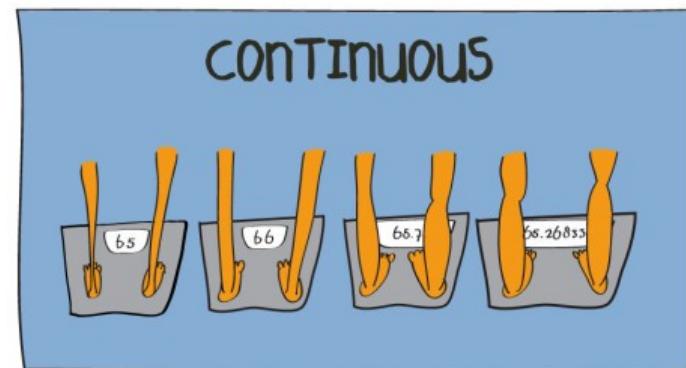
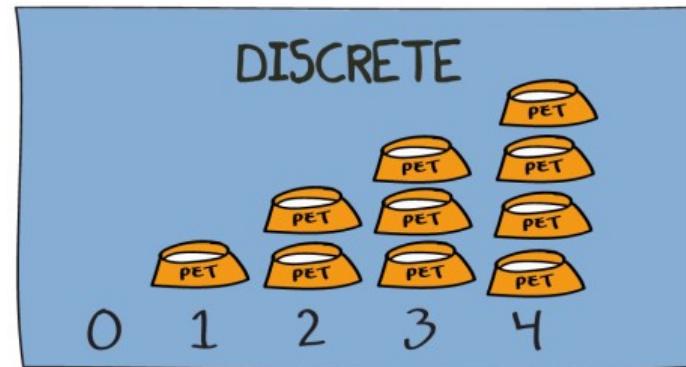
# Quantitative Data

- Quantitative: measured on a numerical scale and can be counted, ordered and measured
- Interval: ordered units that have the same difference
  - Arbitrary zero
  - E.g. Temperature data
- Ratio data: same as interval data but they have an absolute zero
  - Has an absolute zero
  - E.g. Weight, height, length, age...
- Wider range of statistical tests are possible, but most tests have very specific and limiting assumptions about the data

# Types of Data

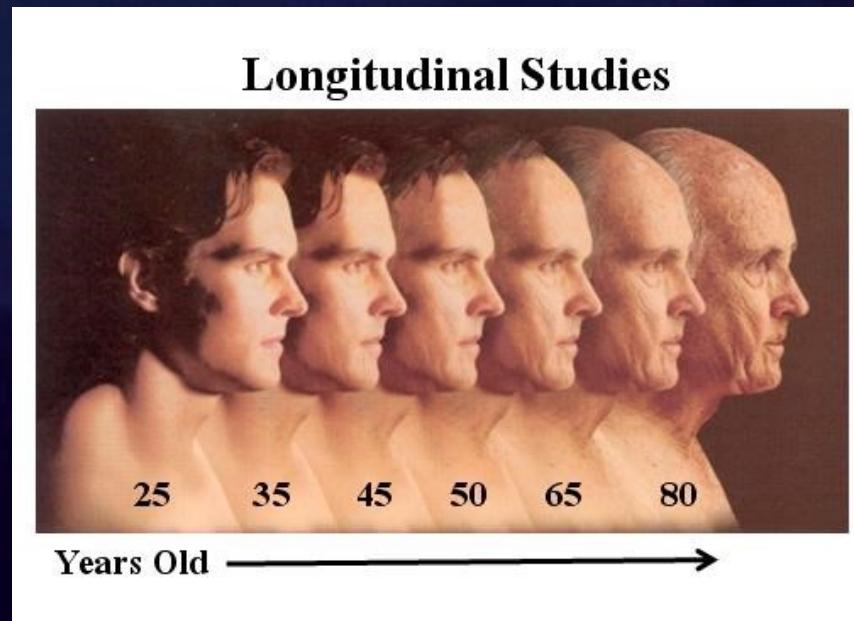
- Discrete variables
  - Values that can be counted (number of children in a family; number of students in a class; income of Professors; 0,1,2,3..)
- Continuous variables
  - These variables can assume an infinite number of values between any two specific values.
  - Obtained by measuring, often fractions and decimals (e.g. temperature).

## Discrete and continuous variables



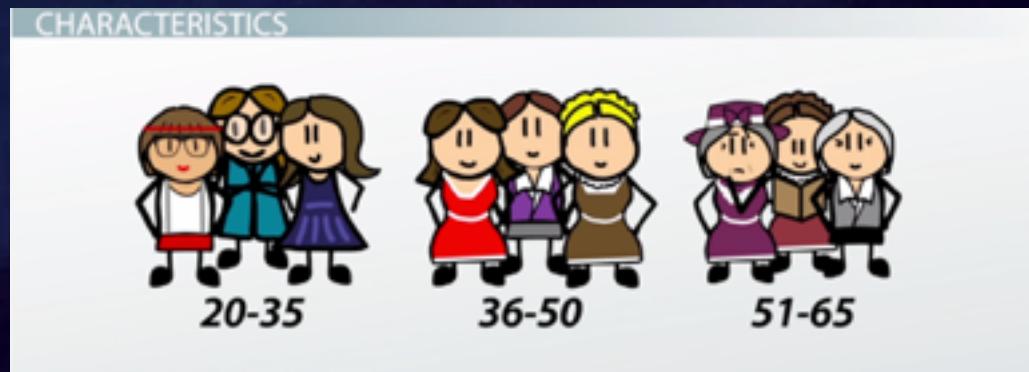
# Data Collection

- *Longitudinal Research*: Collect data at a number of time points
- Cross-sectional Research: Collect data at only one time point



# Data Collection

- *Longitudinal Research*: Collect data at a number of time points
- Cross-sectional Research: Collect data at only one time point



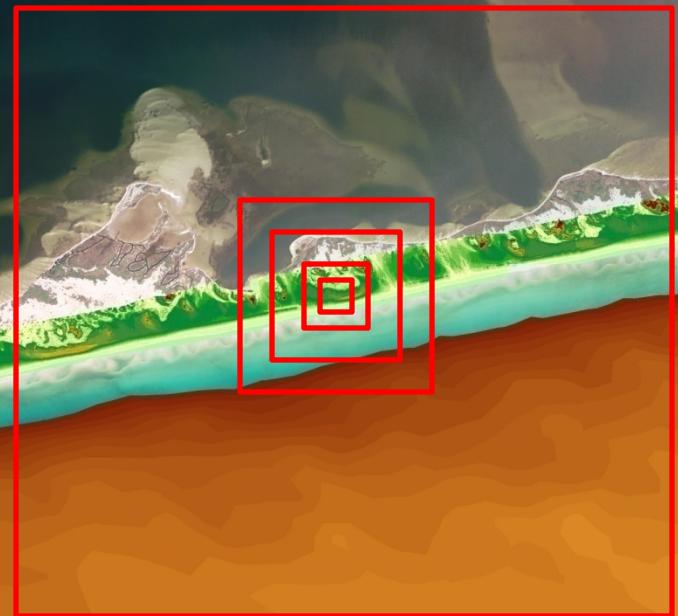
# Limited Resources

Field sampling is not simple and requires consideration of:

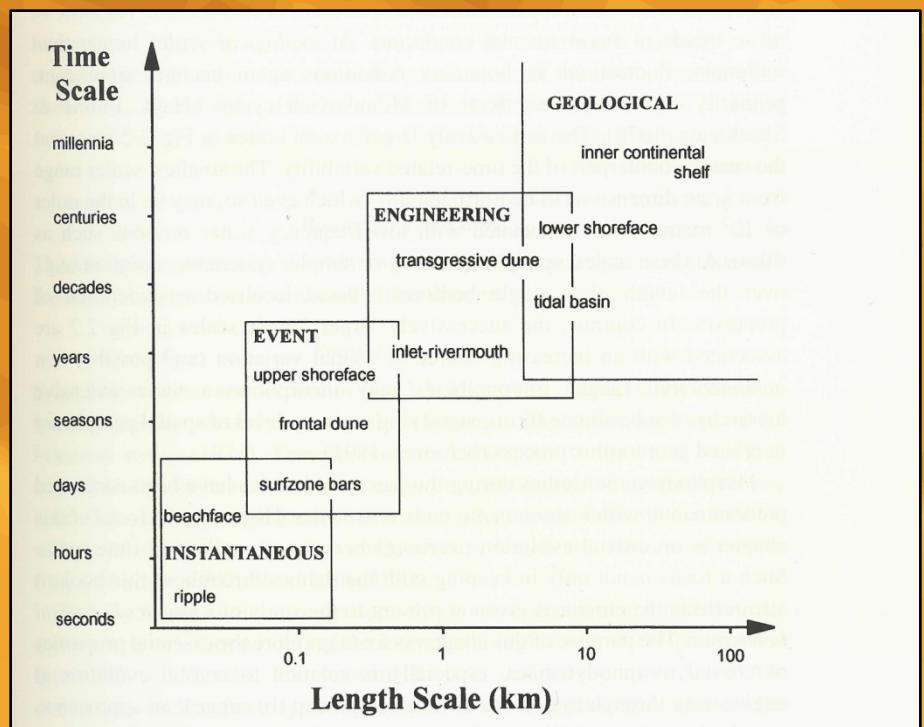
- Legacy effects
- Limited equipment
- Limited time

Consider how you could measure waves and currents within this reef system





# Time and Space Scales



# Space for Time Substitution



# Nature's Rhythm

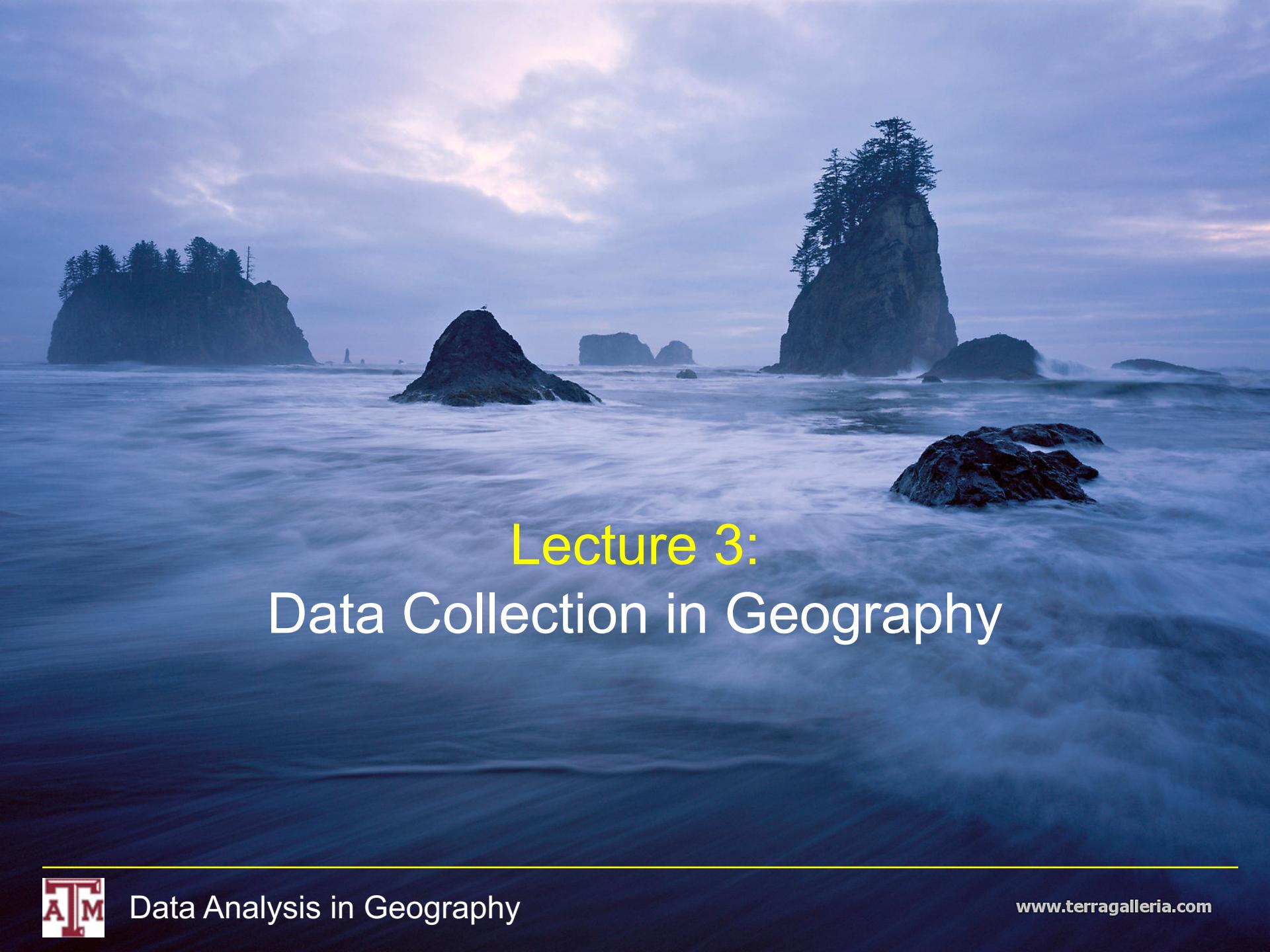
A wide-angle photograph of a natural landscape. In the background, a large, rugged mountain peak is covered with patches of snow. To its right, a massive glacier is visible, its white surface contrasting with the surrounding green and brown terrain. In the foreground, a river flows from the glacier area towards the viewer, its dark blue water reflecting the sky. The banks of the river are rocky and uneven. Beyond the river, there is a field of low-lying vegetation with autumn-colored leaves, ranging from yellow to orange. The sky above is a clear, pale blue with a few wispy clouds.

- Sampling needs to explicitly recognize the natural variability of a phenomenon
  - River discharge
  - Storms on the coast
  - Seasonal variation in vegetation
- Need to sample at least twice the frequency of the variation in order for it to be captured

# Reminders

1. Next time we are going to talk about sampling procedures and then we will be doing some data collection outside in groups. Please bring something to write on (e.g. a notebook or a tablet) and dress for the weather.





# Lecture 3: Data Collection in Geography

