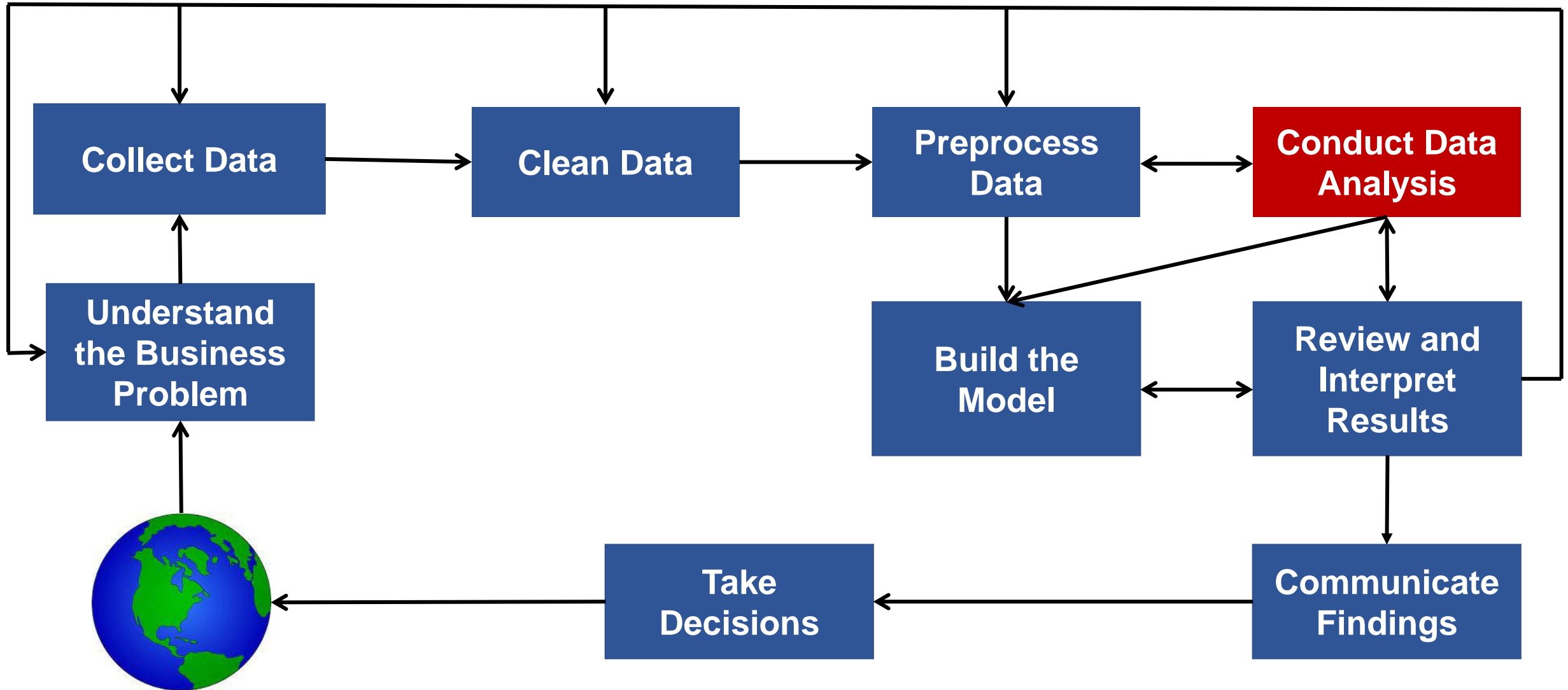


Descriptive Data Analysis

Dr. Sandareka Wickramanayake

sandarekaw@cse.mrt.ac.lk

Data Science Process



Recommended Reading

Chapter 3 – **Art of Data Science**

Chapter 1, 2, 3 – **Head First Statistics**

Data Analysis

- Ask **good** questions.
- Seek answers to those questions.

Descriptive

Seek to summarize characteristics of a set of data.

Exploratory

Seek for patterns, trends, or relationships between variables.

Predictive

Make predictions about future or otherwise unknown events based on data.

Data Analysis

- What is a **good** question?
 - Interesting to the audience
 - Depends on the context and environment in which you are working with the data.
 - Not already been answered
 - Plausible
 - Answerable
 - Relevant data should exist or there should be means to collect the required data.
 - Specific
 - E.g., Is eating a healthier diet better for you? Vs Does eating at least 5 servings per day of fresh fruits and vegetables lead to fewer upper respiratory tract infections (colds)?

Data Analysis

- Ask **good** questions.
- Seek answers to those questions.

Descriptive

Seek to summarize a characteristic of a set of data.

Exploratory

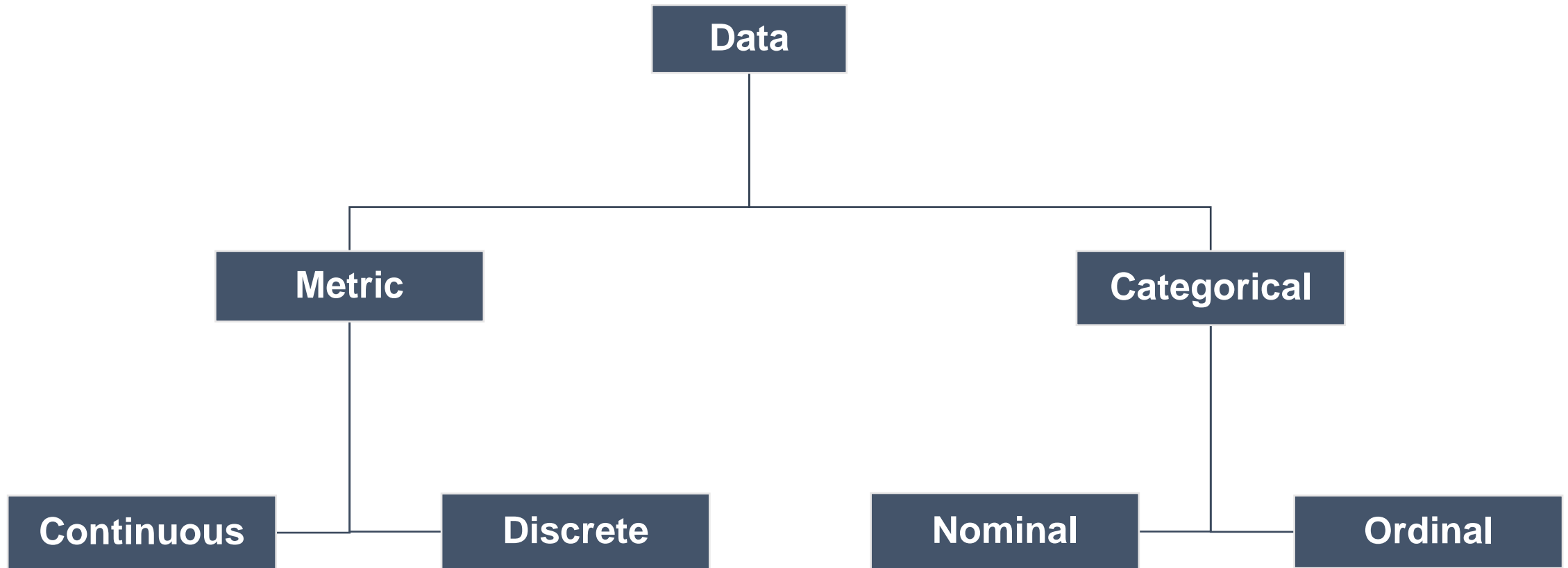
Seek for patterns, trends, or relationships between variables.

Predictive

Make predictions about future or otherwise unknown events based on data.

Data Types (Recap)

- Four fundamental data types.



Data Types (Recap)

- Nominal (Categorical)

- Attribute: Gender Values: {Male, Female}

Participant ID	Gender	Age	Height (m)	Hair Color	Seatbelt
1	Male	18	1.8	Black	2_Rarely
2	Female	14	1.57	Red	2_Rarely
3	Male	17	1.73	Black	3_Sometimes
4	Male	22	1.7	Black	5_Always
5	Male	23	1.81	Black	3_Sometimes
6	Female	21	1.63	Brown	3_Sometimes
7	Male	23	1.7	Red	5_Always
8	Female	21	1.55	Brown	5_Always
9	Female	19	1.62	Red	3_Sometimes
10	Female	20	1.6	Black	5_Always
11	Male	18	1.71	Brown	2_Rarely
12	Male	25	1.65	Brown	2_Rarely
13	Female	13	1.56	Brown	2_Rarely
14	Male	17	1.62	Red	4_Mosttimes

Special Case: If there are only two categories (e.g., Male/Female, True/False), we call it a **Boolean attribute**

Data Types (Recap)

- Ordinal (Categorical)

- Attribute: Where_A_Seatbelt
Rarely, Never}

Values: {Most_times, Sometimes,

Participant ID	Gender	Age	Height (m)	Hair Color	Seatbelt
1	Male	18	1.8	Black	2_Rarely
2	Female	14	1.57	Red	2_Rarely
3	Male	17	1.73	Black	3_Sometimes
4	Male	22	1.7	Black	5_Always
5	Male	23	1.81	Black	3_Sometimes
6	Female	21	1.63	Brown	3_Sometimes
7	Male	23	1.7	Red	5_Always
8	Female	21	1.55	Brown	5_Always
9	Female	19	1.62	Red	3_Sometimes
10	Female	20	1.6	Black	5_Always
11	Male	18	1.71	Brown	2_Rarely
12	Male	25	1.65	Brown	2_Rarely
13	Female	13	1.56	Brown	2_Rarely
14	Male	17	1.62	Red	4_Mosttimes

Basic arithmetic is not appropriate.

Data Types (Recap)

- Discrete (Metric)

- Attribute: Age

Values: {1, 2, 3, 4, 5, 6, ..., 100}

Participant ID	Gender	Age	Height (m)	Hair Color	Seatbelt	
1	Male	18	1.8	Black	2_Rarely	
2	Female	14	1.57	Red	2_Rarely	
3	Male	17	1.73	Black	3_Sometimes	
4	Male	22	1.7	Black	5_Always	
5	Male	23	1.81	Black	3_Sometimes	
6	Female	21	1.63	Brown	3_Sometimes	
7	Male	23	1.7	Red	5_Always	
8	Female	21	1.55	Brown	5_Always	
9	Female	19	1.62	Red	3_Sometimes	
10	Female	20	1.6	Black	5_Always	
11	Male	18	1.71	Brown	2_Rarely	
12	Male	25	1.65	Brown	2_Rarely	
13	Female	13	1.56	Brown	2_Rarely	
14	Male	17	1.62	Red	4_Mosttimes	

Data Types (Recap)

- Continuous (Metric)

- Attribute: Height Values: {1, 1.2, 1.3, ..., 3}

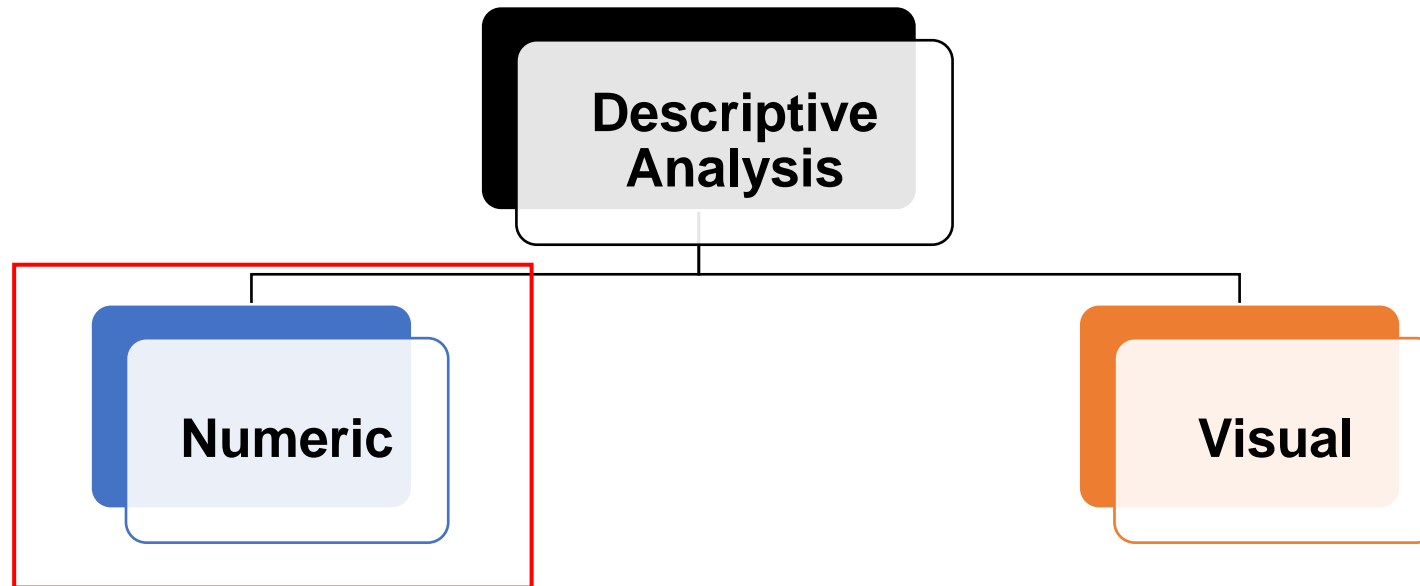
Participant ID	Gender	Age	Height (m)	Hair Color	Seatbelt	
1	Male	18	1.8	Black	2_Rarely	
2	Female	14	1.57	Red	2_Rarely	
3	Male	17	1.73	Black	3_Sometimes	
4	Male	22	1.7	Black	5_Always	
5	Male	23	1.81	Black	3_Sometimes	
6	Female	21	1.63	Brown	3_Sometimes	
7	Male	23	1.7	Red	5_Always	
8	Female	21	1.55	Brown	5_Always	
9	Female	19	1.62	Red	3_Sometimes	
10	Female	20	1.6	Black	5_Always	
11	Male	18	1.71	Brown	2_Rarely	
12	Male	25	1.65	Brown	2_Rarely	
13	Female	13	1.56	Brown	2_Rarely	
14	Male	17	1.62	Red	4_Mosttimes	

Descriptive Analysis

- Summarizes the properties of the dataset.
- Example questions that can be answered using descriptive analysis are
 - *Determine the proportion of males.*
 - *What is the mean age of the participants?*
 - *What percentage of participants “rarely” uses the seatbelt?*
 - Etc ...
- *An important consideration is the type of data!*
 - Step 1: Use data type classification tree.
 - Step 2: Apply appropriate descriptive methods.

Descriptive Analysis

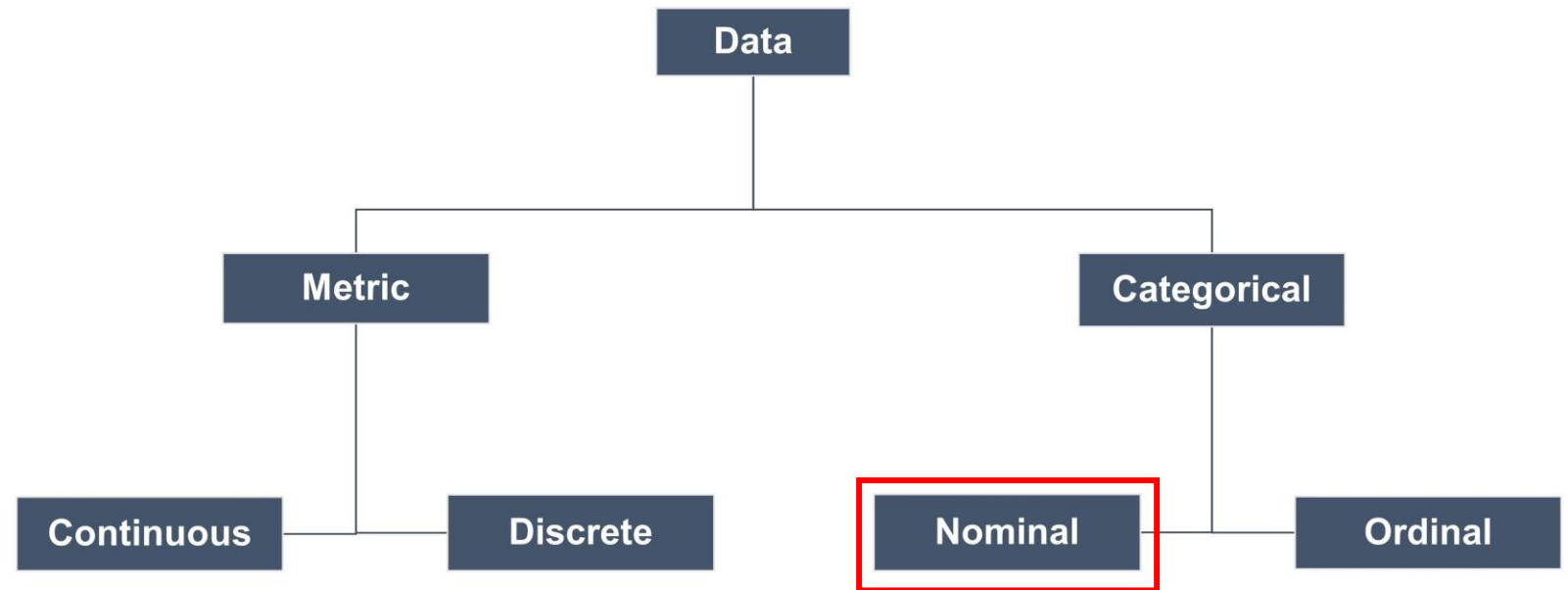
- Two main approaches: Numeric and Visual descriptions.



Descriptive Analysis : Numeric

- **Nominal Data**

- Frequency
- Normalized Frequency
- ~~Cumulative Frequency~~
- ~~Normalized Cumulative Frequency~~
- Mode
- ~~Median~~
- ~~Mean~~
- ~~Range~~
- ~~Spread~~



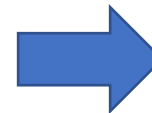
Descriptive Analysis: Numeric

- **Frequency**

- Count the number of occurrences.
- Order of name appearance is not relevant but to improve comprehension we should sort the frequency!
- *Ex. question – How many participants have black hair?*

Participant ID	Hair Color
1	Black
2	Red
3	Black
4	Black
5	Black
6	Brown
7	Red
8	Brown
9	Red
10	Black

Hair Color	Frequency
Brown	2
Red	3
Black	5



Hair Color	Frequency
Black	5
Red	3
Brown	2

Descriptive Analysis: Numeric

- **Normalized Frequency**

- Count the number of occurrences.
- Then normalized to the total.
- Also known as *relative* or *percentage frequency*
- *Ex. question – What is the proportion of participants having black hair?*

Participant ID	Hair Color
1	Black
2	Red
3	Black
4	Black
5	Black
6	Brown
7	Red
8	Brown
9	Red
10	Black

Hair Color	Frequency	Normalized Frequency (%)
Black	5	50
Red	3	30
Brown	2	20

Descriptive Analysis: Numeric

- **Mode**

- Label that has the highest frequency.
- What if we have more than one label attaining the highest frequency?
 - Multiple modes
 - Report them all
- *Ex. question – What is the most popular hair color?*

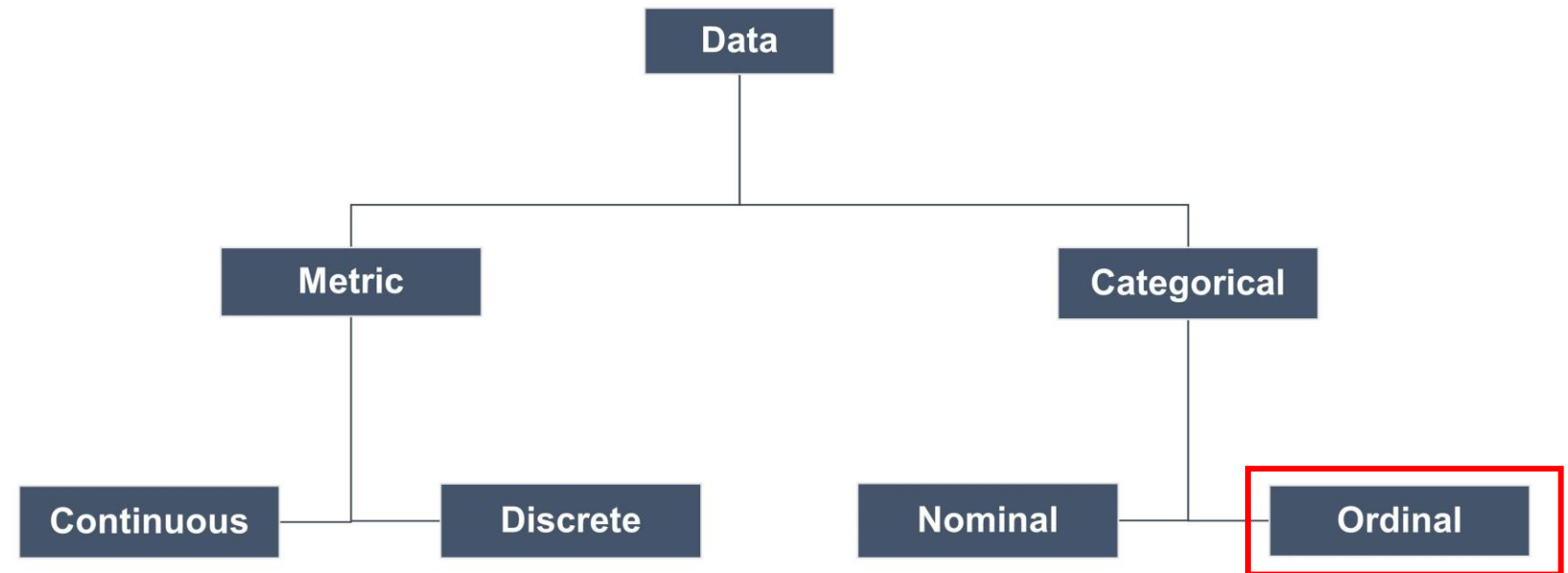
Participant ID	Hair Color
1	Black
2	Red
3	Black
4	Black
5	Black
6	Brown
7	Red
8	Brown
9	Red
10	Black

Hair Color	Frequency
Black	5
Red	3
Brown	2

Descriptive Analysis: Numeric

- **Ordinal Data**

- Frequency
- Normalized Frequency
- Cumulative Frequency
- Normalized Cumulative Frequency
- Mode
- Median
- ~~Mean~~
- ~~Range~~
- ~~Spread~~



Descriptive Analysis: Numeric

• Cumulative Frequency

- Accumulated count.
- Values must be ordered.
- *Ex. Question - How many participants **did not always** wear the seatbelt?*

Participant ID	Seatbelt
1	2_Rarely
2	2_Rarely
3	3_Sometimes
4	5_Always
5	3_Sometimes
6	3_Sometimes
7	5_Always
8	5_Always
9	3_Sometimes
10	5_Always
11	2_Rarely
12	2_Rarely
13	2_Rarely
14	4_Mosttimes

Values
must be
sorted or
ordered

Seatbelt	Frequency	Cumulative Frequency
Never	0	0
Rarely	5	$5=5+0$
Sometimes	4	$9=4+5$
Mosttimes	1	$10=1+9$
Always	4	$14=4+9$

Descriptive Analysis: Numeric

- **Normalized Cumulative Frequency**

- Accumulated count.
- Values must be ordered.
- *Ex. Question – What is the proportion of participants who **did not always** wear the seatbelt?*

Participant ID	Seatbelt	
1	2_Rarely	
2	2_Rarely	
3	3_Sometimes	
4	5_Always	
5	3_Sometimes	
6	3_Sometimes	
7	5_Always	
8	5_Always	
9	3_Sometimes	
10	5_Always	
11	2_Rarely	
12	2_Rarely	
13	2_Rarely	
14	4_Mosttimes	

Values
must be
sorted or
ordered

Seatbelt	Normalized Frequency	Normalized Cumulative Frequency
Never	0	0
Rarely	35.7	$35.7=35.7+0$
Sometimes	28.6	$64.3=35.7+28.6$
Mosttimes	7.1	$71.4=64.3+7.1$
Always	28.6	$100=71.4+28.6$

Descriptive Analysis: Numeric

- **Median**

- We can calculate the median because ordinal data can be ordered.
- Arrange data in increasing order.
- Median is the value in the middle position.
 - If you have n numbers, the middle position = $\frac{n+1}{2}$.
- If you have an odd number of values
 - Median is the one in the middle position.
- If you have an even number of values
 - Median = $\frac{\text{Summation of the two middle values}}{2}$
 - Two middle values are on either side of the point $\frac{n+1}{2}$.

Descriptive Analysis: Numeric

- Median

Participant ID	Seatbelt
1	2_Rarely
2	2_Rarely
3	3_Sometimes
4	5_Always
5	3_Sometimes
6	3_Sometimes
7	5_Always
8	5_Always
9	3_Sometimes
10	5_Always
11	2_Rarely
12	2_Rarely
13	2_Rarely
14	4_Mosttimes

Seatbelt	Frequency
Never	0
Mosttimes	1
Sometimes	4
Always	4
Rarely	5

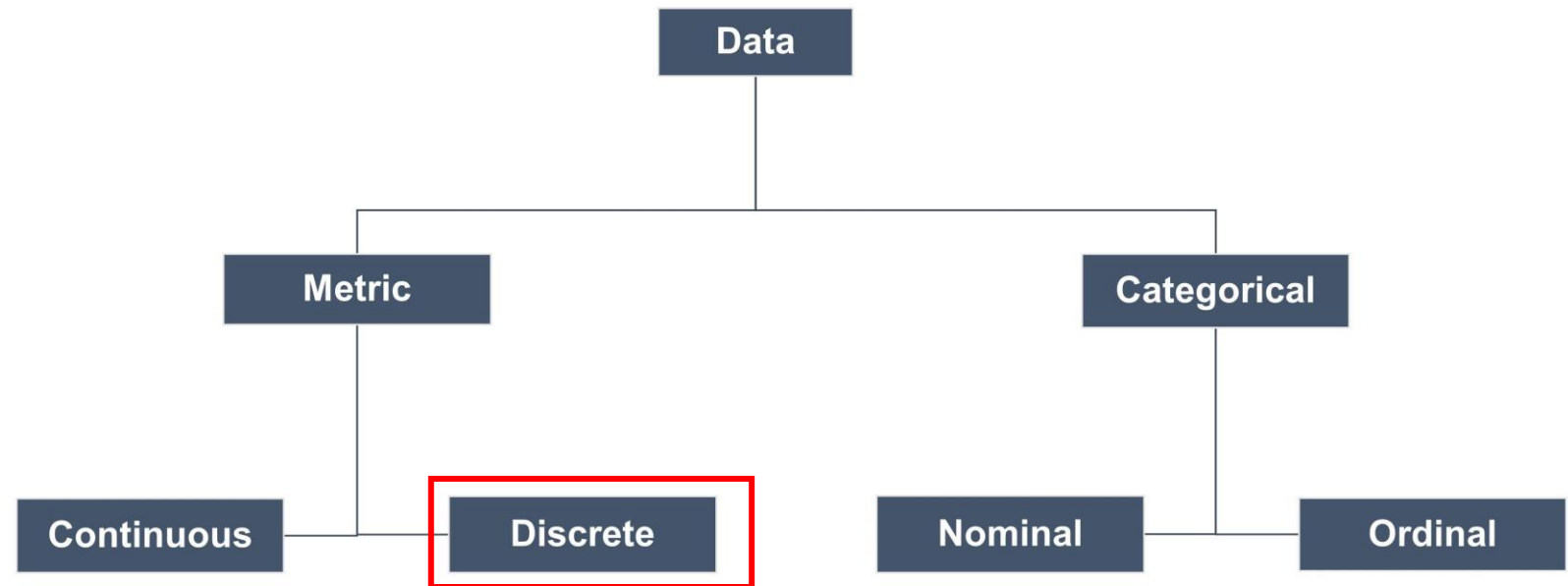
Middle position – $7.5 = (14+1)/2$

Median = “Always”

Descriptive Analysis: Numeric

- **Discrete Data**

- Frequency
- Normalized Frequency
- Cumulative Frequency
- Normalized Cumulative Frequency
- Mode
- Median
- Mean
- Range
- Spread



Descriptive Analysis: Numeric

- **Mean**

- One of the most frequently used statistics.
- The most common measure of the location of asset of points.

Participant ID	Age
1	18
2	14
3	17
4	22
5	23
6	21
7	23
8	21
9	19
10	20
11	18
12	25
13	13
14	17

$$\mu = \frac{\sum x}{n}$$

$$19.4 =$$

$$\frac{18 + 14 + 17 + 22 + 23 + 21 + 23 + 21 + 19 + 20 + 18 + 25 + 13 + 17}{14}$$

Descriptive Analysis: Numeric

- **Mean**

- One of the most frequently used statistics.
- The most common measure of the location of asset of points.
- **Very sensitive to outliers.**

Participant ID	Age
1	18
2	14
3	17
4	22
5	23
6	21
7	23
8	21
9	19
10	20
11	18
12	25
13	13
14	17

$$\mu = \frac{\sum x}{n}$$

26.7 =

$$\frac{18 + 14 + 17 + 22 + 23 + 21 + 23 + 21 + 19 + 20 + 18 + \mathbf{75} + 13 + \mathbf{70}}{14}$$

Descriptive Analysis: Numeric

- **Range**

- Difference between the *max* and *min*.

Participant ID	Age
1	18
2	14
3	17
4	22
5	23
6	21
7	23
8	21
9	19
10	20
11	18
12	25
13	13
14	17

$$Range = 12 = 25 - 13$$

Descriptive Analysis: Numeric

- **Variance = (Standard Deviation)²**

- Most common measures of the spread of a set of points.
- **Very sensitive to outliers.**

Participant ID	Age
1	18
2	14
3	17
4	22
5	23
6	21
7	23
8	21
9	19
10	20
11	18
12	25
13	13
14	17

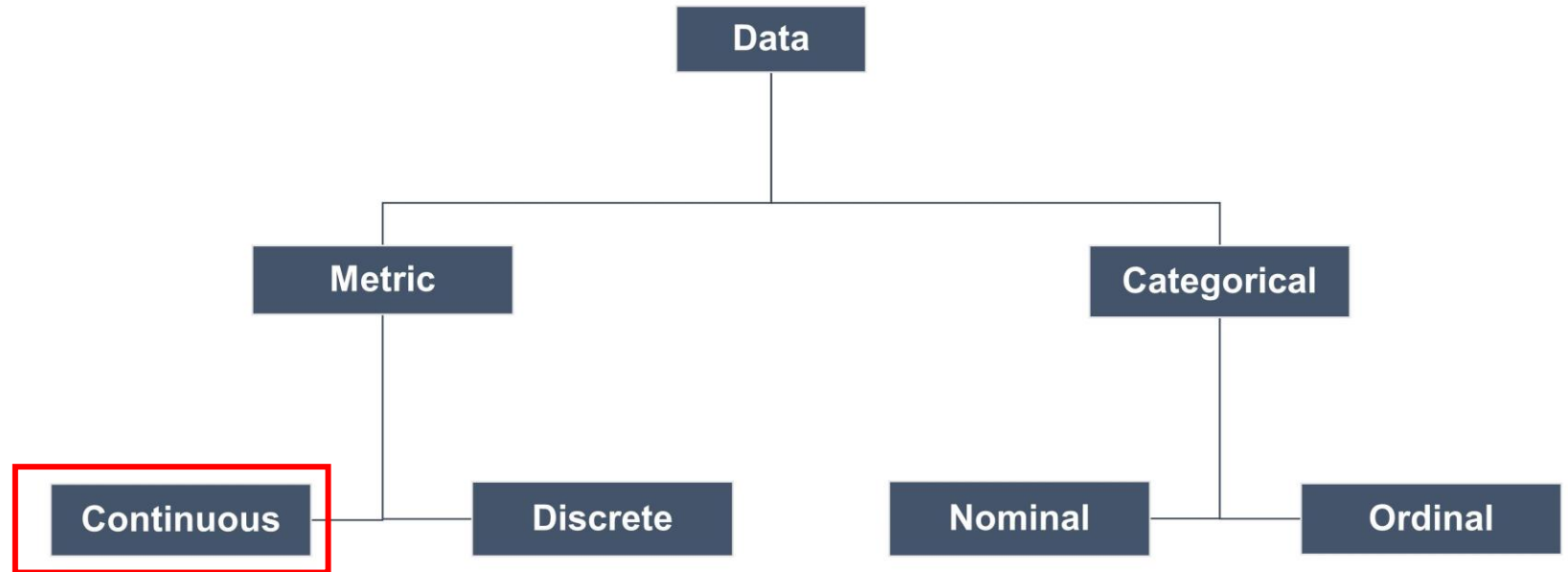
$$\text{sample variance} = s^2 = \frac{1}{n - 1} \sum (x - \mu)^2$$

$$\text{population variance} = s^2 = \frac{1}{n} \sum (x - \mu)^2$$

Descriptive Analysis: Numeric

- **Continuous Data**

- Grouped Representation
 - Frequency
 - Normalized Frequency
 - Cumulative Frequency
 - Normalized Cumulative Frequency
- Mode
- Median
- Mean
- Range
- Spread



Descriptive Analysis: Numeric

- Grouped Representation
 - Divide data into consecutive intervals of values.
 - Turn continuous data to ordinal data.

Participant ID	Height (m)
1	1.8
2	1.57
3	1.73
4	1.7
5	1.81
6	1.63
7	1.7
8	1.55
9	1.62
10	1.6
11	1.71
12	1.65
13	1.56

Height	Frequency	Normalized Frequency (%)	Cumulative Frequency	Cumulative Normalised Frequency (%)
1.5 – 1.6	3	23.1	3	23.1
1.6 – 1.7	6	46.1	9	69.2
1.7 – 1.8	3	23.1	12	92.3
1.8 – 1.9	1	7.7	13	100
> 1.9	0	0	13	100

Descriptive Analysis: Numeric

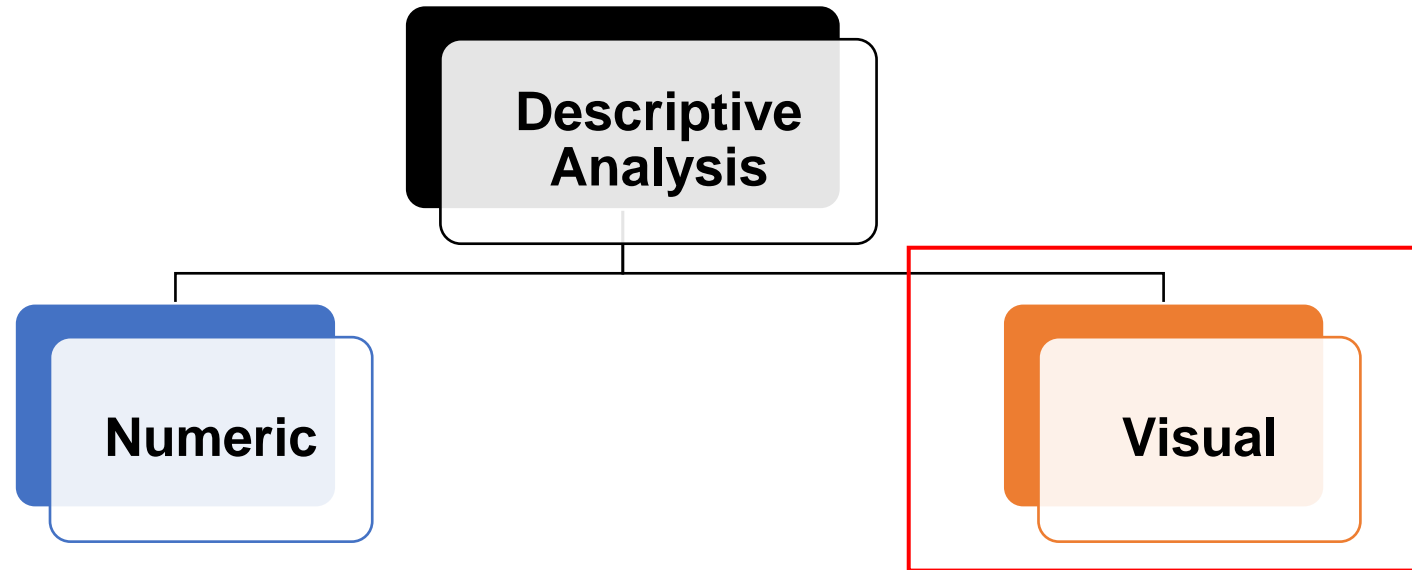
- **Five-number Summary**
 - Min
 - Lower Quartile (Q1) – 25th Percentile
 - Median (Q2) – 50th Percentile
 - Upper Quartile – 75th Percentile
 - Max

Descriptive Analysis: Numeric

Attribute Type	Categorical		Metric	
	Nominal	Ordinal	Discrete	Continuous
Frequency	Yes	Yes	Yes	Under grouped representation
Normalized Frequency	Yes	Yes	Yes	Under grouped representation
Cumulative Frequency	No	Yes	Yes	Under grouped representation
Normalized Cumulative Frequency	No	Yes	Yes	Under grouped representation
Mode	Yes	Yes	Yes	No
Mean	No	No	Yes	Yes
Median	No	Yes	Yes	Yes
Range	No	No	Yes	Yes
Spread	No	No	Yes	Yes
Five Number Summary	No	No	Yes	Yes

Descriptive Analysis

- Two main approaches: Numeric and Visual descriptions.



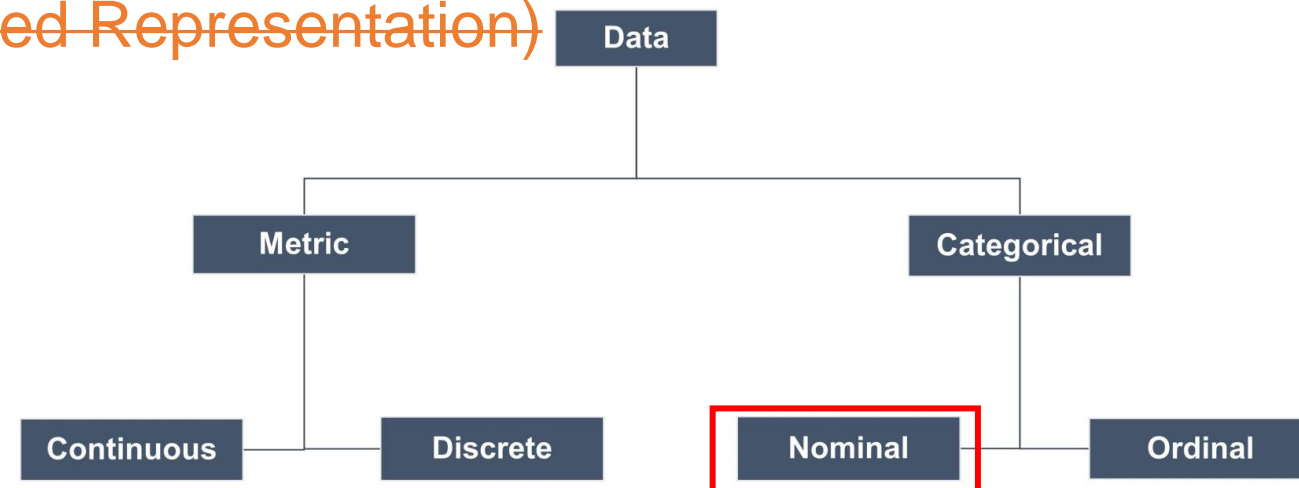
Descriptive Analysis: Visual

- Visualization
 - Conversion of data into a **visual** format so that characteristics of the data and relationships among data items or attributes can be analysed or reported.
- One of the most powerful and appealing techniques for data exploration.
 - Humans have a well-developed ability to analyse a large amount of data presented visually.
 - Can detect **general** patterns and trends.
 - Can detect outliers and **unusual** patterns.

Descriptive Analysis: Visual

- **Nominal Data**

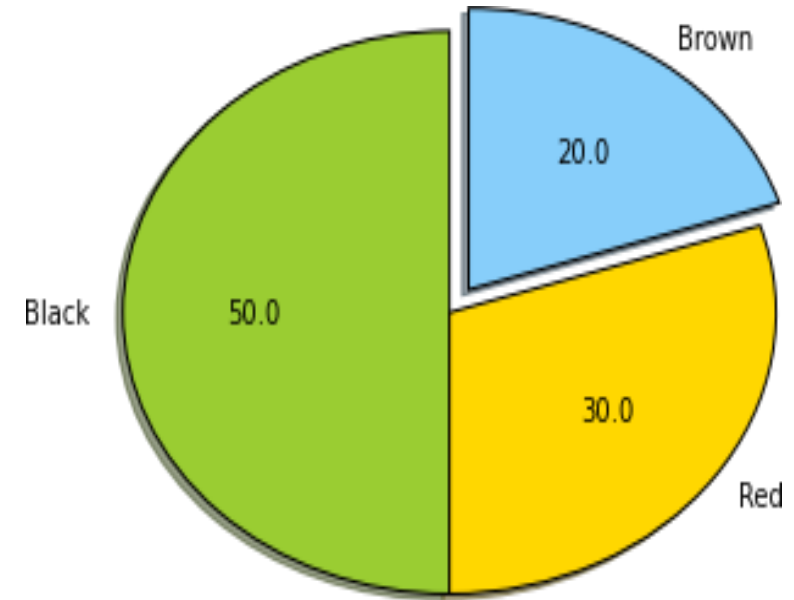
- Pie chart (for a small number of labels)
- Tag Cloud (for a large number of labels)
- Bar Chart
- Clustered/Stacked Bar Chart
- ~~Step Chart~~
- ~~Box-Plot~~
- ~~Histograms (Grouped Representation)~~
- ~~Cumulative Histograms (Grouped Representation)~~



Descriptive Analysis: Visual

- **Pie Chart**

- Divided into sectors.
- Describes the percentage for each label.



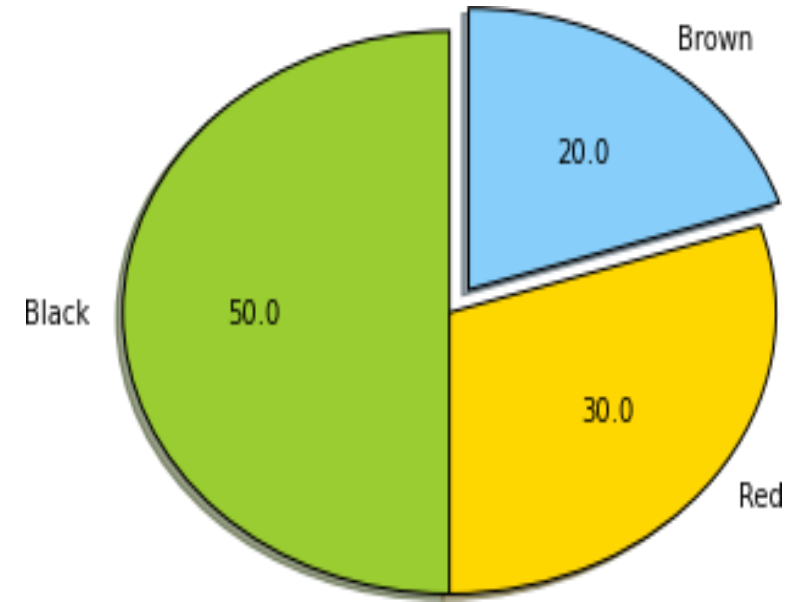
Participant ID	Hair Color
1	Black
2	Red
3	Black
4	Black
5	Black
6	Brown
7	Red
8	Brown
9	Red
10	Black

Hair Color	Frequency	Normalized Frequency (%)
Black	5	50
Red	3	30
Brown	2	20

Descriptive Analysis: Visual

- **Pie Chart**

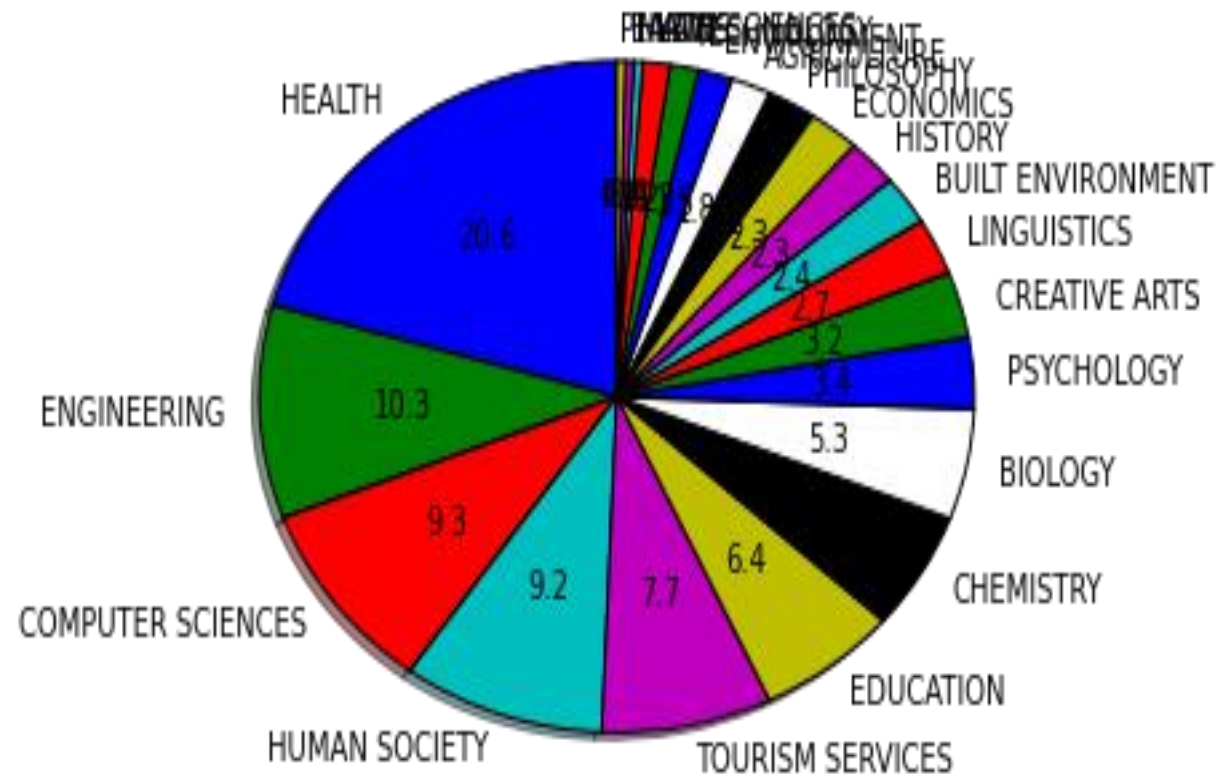
- Divided into sectors.
- Describes the percentage for each label.
- Visually appeal.
- Gain understanding in the distribution.
- Only suitable for small number of labels.



Descriptive Analysis: Visual

- Pie Chart

Field of Research	Percentage
HEALTH	20.61%
ENGINEERING	10.25%
COMPUTER SCIENCES	9.32%
HUMAN SOCIETY	9.15%
TOURISM & SERVICES	7.70%
EDUCATION	6.36%
CHEMISTRY	5.70%
BIOLOGY	5.33%
PSYCHOLOGY	3.43%
CREATIVE ARTS	3.17%
LINGUISTICS	2.73%
BUILT ENVIRONMENT	2.39%
HISTORY	2.31%
ECONOMICS	2.30%
PHILOSOPHY	2.22%
AGRICULTURE	1.79%
ENVIRONMENT	1.55%
TECHNOLOGY	1.29%
LAW	1.23%
MATHS	0.40%
EARTH SCIENCES	0.39%
PHYSICS	0.38%



- Hard to read
- Visually ugly

Descriptive Analysis: Visual

• Tag Cloud

Field of Research	Percentage
HEALTH	20.61%
ENGINEERING	10.25%
COMPUTER SCIENCES	9.32%
HUMAN SOCIETY	9.15%
TOURISM & SERVICES	7.70%
EDUCATION	6.36%
CHEMISTRY	5.70%
BIOLOGY	5.33%
PSYCHOLOGY	3.43%
CREATIVE ARTS	3.17%
LINGUISTICS	2.73%
BUILT ENVIRONMENT	2.39%
HISTORY	2.31%
ECONOMICS	2.30%
PHILOSOPHY	2.22%
AGRICULTURE	1.79%
ENVIRONMENT	1.55%
TECHNOLOGY	1.29%
LAW	1.23%
MATHS	0.40%
EARTH SCIENCES	0.39%
PHYSICS	0.38%



<http://wordle.net>

- The size of the label is proportional to the volume.
- Visually appealing.
- **Limitation: Don't know precisely the proportion.**

Descriptive Analysis: Visual

- **Bar Chart**

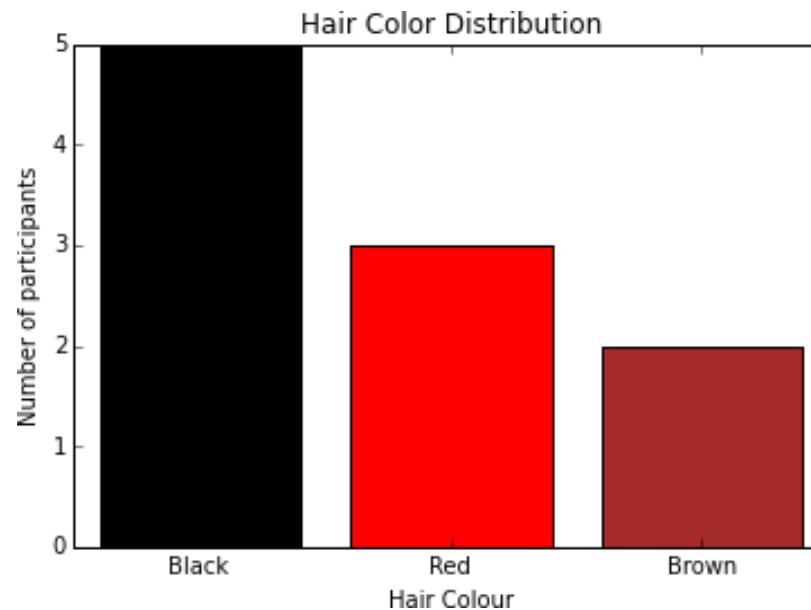
- Labels on one axis and count/frequency/normalized frequency on the other axis.
- Three common types:
 - **Simple bar chart:** to describe a single (nominal) variable.
 - **Clustered bar chart:** to describe two or more variables.
 - **Stacked bar chart:** to describe two or more variables.

Descriptive Analysis: Visual

- **Bar Chart**

- **Simple bar chart:** to describe a single (nominal) variable.
 - Bars should have the same width.
 - There are equal spaces among bars. – To emphasize the categorical nature of the data.

Participant ID	Hair Color
1	Black
2	Red
3	Black
4	Black
5	Black
6	Brown
7	Red
8	Brown
9	Red
10	Black



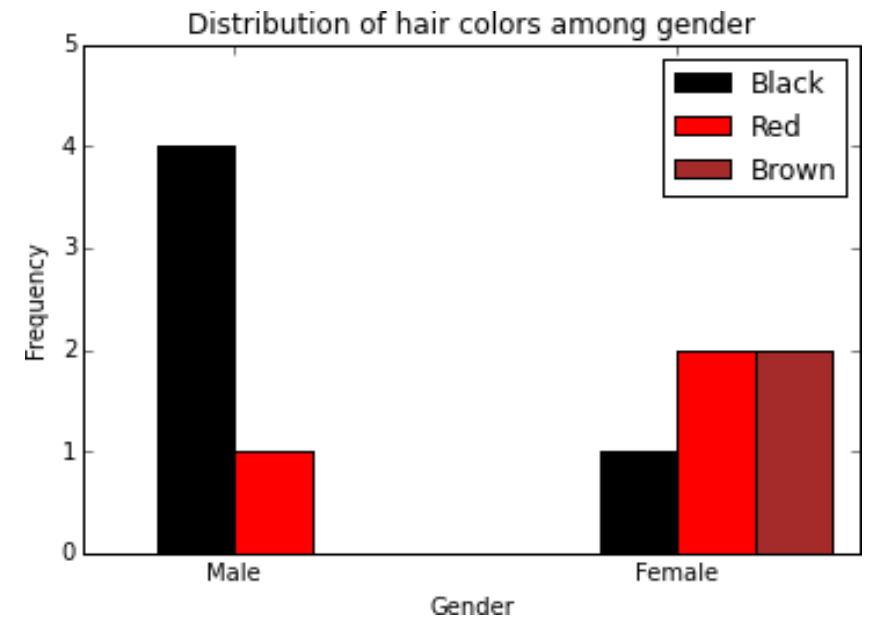
Descriptive Analysis: Visual

- **Bar Chart**

- **Clustered bar chart:** to describe two or more variables.
 - Visualize the statistics of one variable in relation to another variable.
 - Ex. Question – What is the distribution of hair color among female participants?

Participant ID	Gender	Hair Color
1	Male	Black
2	Female	Red
3	Male	Black
4	Male	Black
5	Male	Black
6	Female	Brown
7	Male	Red
8	Female	Brown
9	Female	Red
10	Female	Black

Gender			
Hair Color		Male	Female
	Black	4	1
	Red	1	2
	Brown	0	2



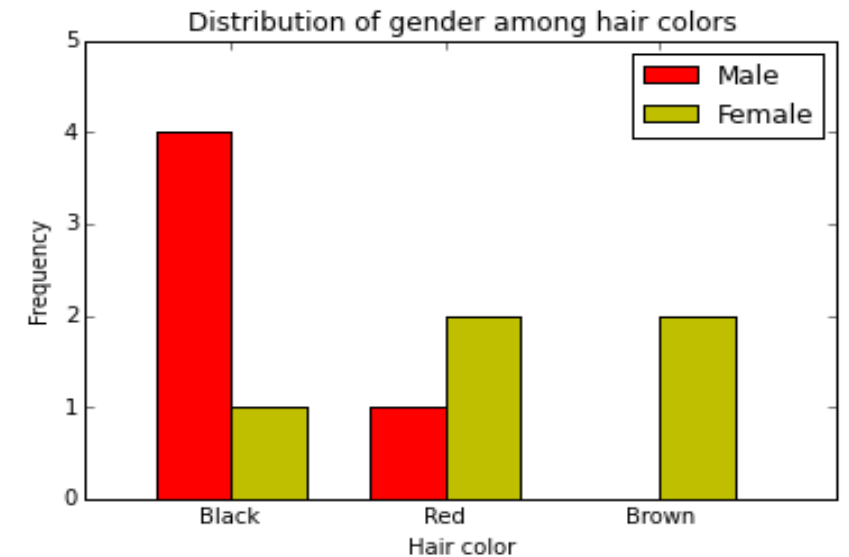
Descriptive Analysis: Visual

- **Bar Chart**

- **Clustered bar chart:** to describe two or more variables.
 - Visualize the statistics of one variable in relation to another variable.
 - Ex. Question – What is the distribution of gender for those having black hair?

Participant ID	Gender	Hair Color
1	Male	Black
2	Female	Red
3	Male	Black
4	Male	Black
5	Male	Black
6	Female	Brown
7	Male	Red
8	Female	Brown
9	Female	Red
10	Female	Black

Gender			
Hair Color		Male	Female
	Black	4	1
	Red	1	2
	Brown	0	2

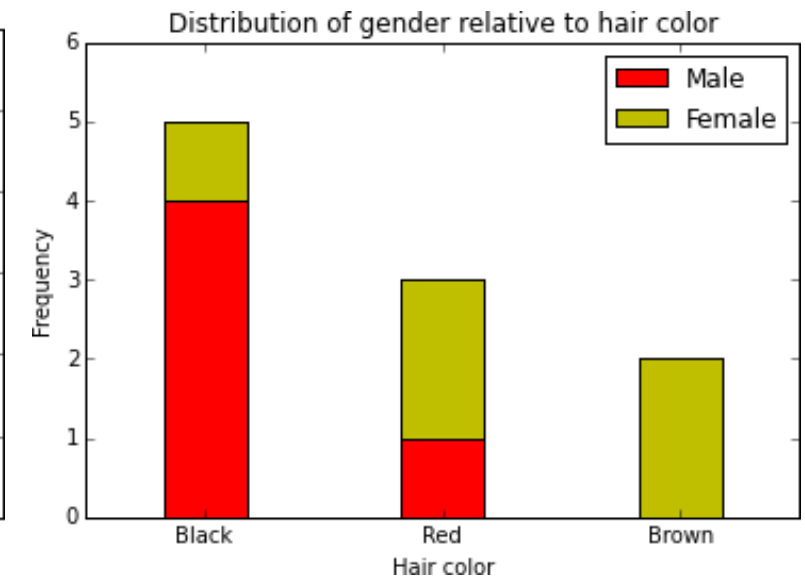
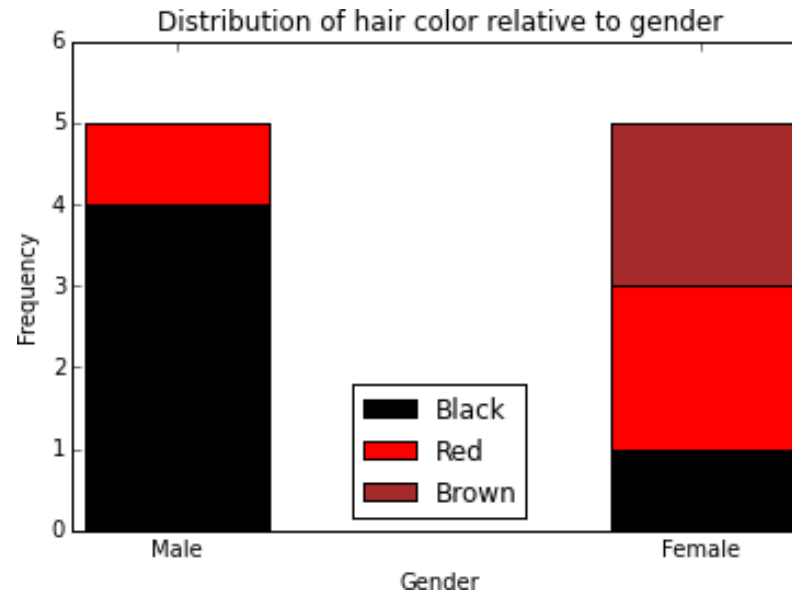


Descriptive Analysis: Visual

- **Bar Chart**

- **Stacked bar chart:** to describe two or more variables.
 - Visualize the statistics of one variable in relation to another variable.
 - Ex. Question – What is the distribution of hair color among female participants?
 - Ex. Question – What is the distribution of gender for those having black hair?
 - Useful when have higher number of values for one variable

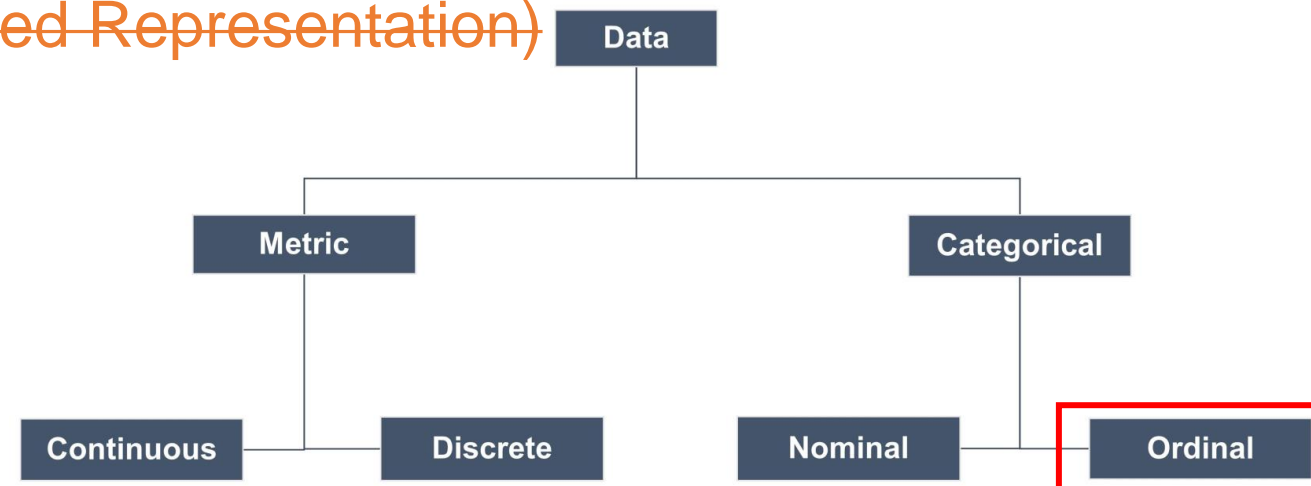
Gender			
Hair Color		Male	Female
	Black	4	1
	Red	1	2
	Brown	0	2



Descriptive Analysis: Visual

• Ordinal Data

- Pie chart (for a small number of labels)
- Tag Cloud (for a large number of labels)
- Bar Chart
- Clustered/Stacked Bar Chart
- Step Chart
- ~~Box-Plot~~
- ~~Histograms (Grouped Representation)~~
- ~~Cumulative Histograms (Grouped Representation)~~

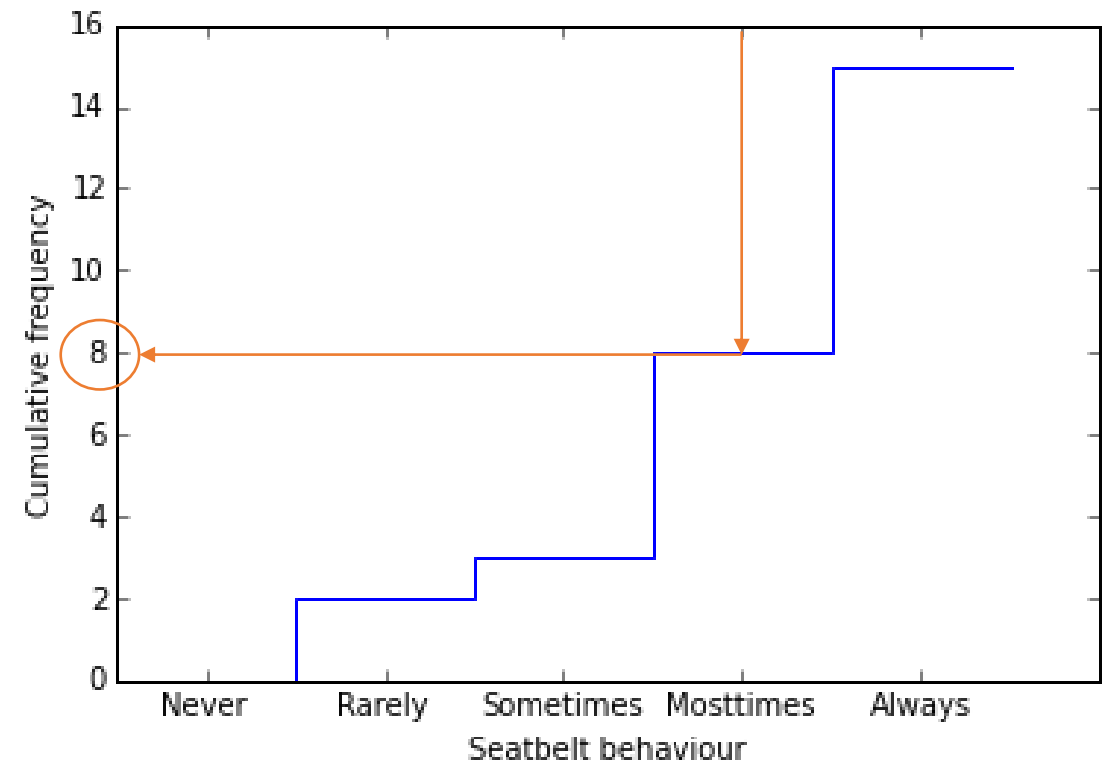


Descriptive Analysis: Visual

- **Step Chart**

- To visualize a **cumulative frequency** or cumulative normalized frequency.
- Ex. Question - **How many** participants **did not** always wear the seatbelt?

Seatbelt	Frequency	Cumulative Frequency
Never	0	0
Rarely	2	2
Sometimes	1	3
Mostimes	5	8
Always	7	15

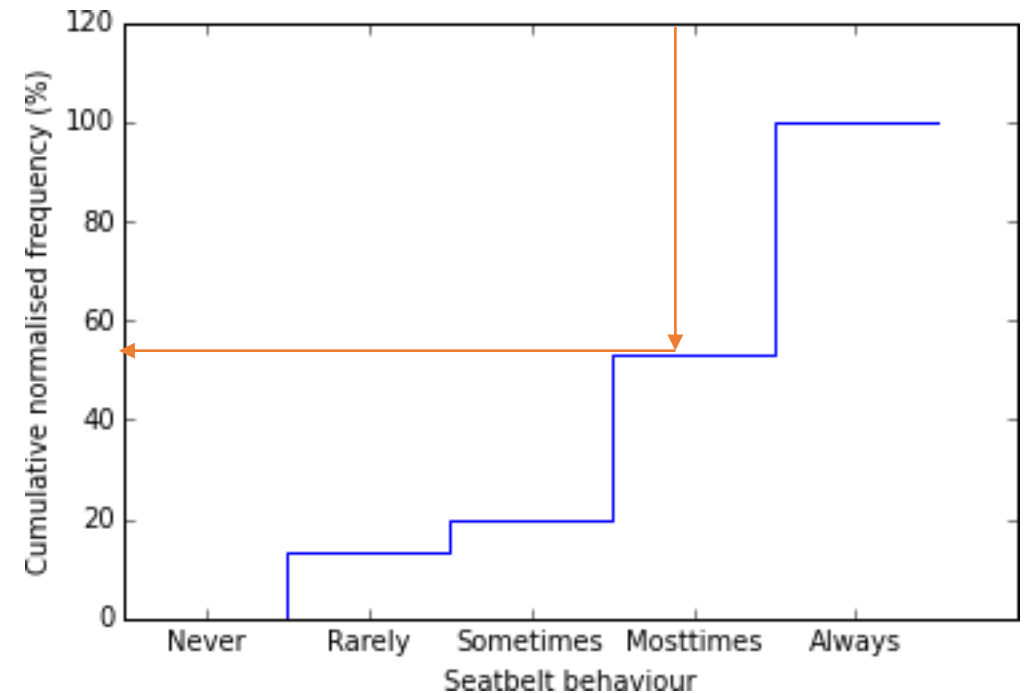


Descriptive Analysis: Visual

- **Step Chart**

- To visualize a cumulative frequency or **cumulative normalized frequency**.
- Ex. Question – **What proportion of participants did not** always wear the seatbelt?

Seatbelt	Frequency	Cumulative Normalized Frequency (%)
Never	0	0
Rarely	2	13.3
Sometimes	1	20.0
Mostimes	5	53.3
Always	7	100

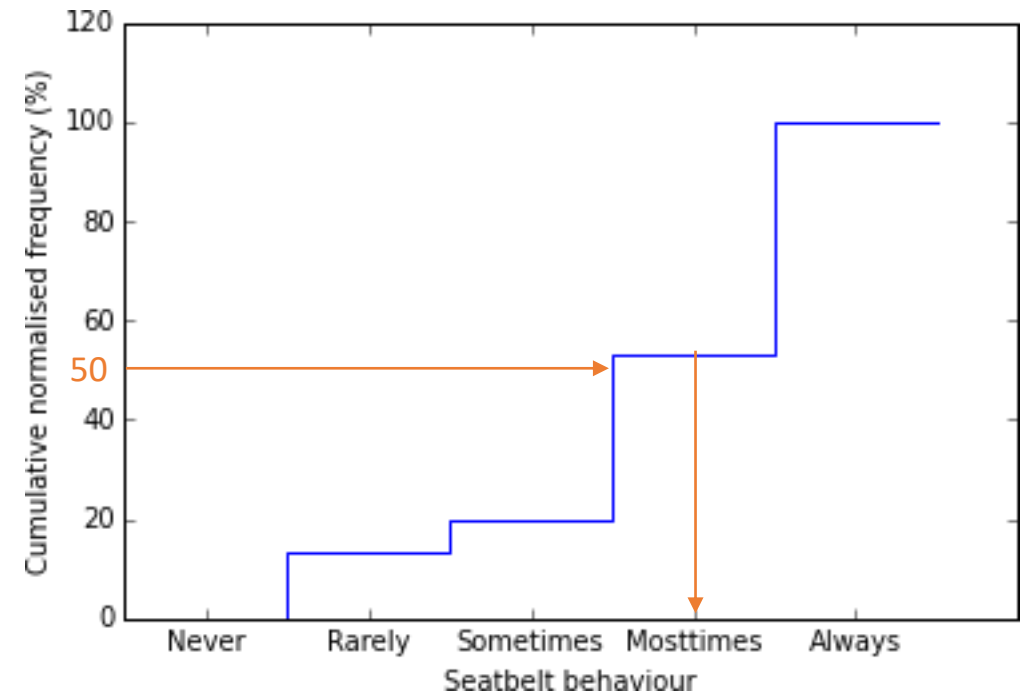


Descriptive Analysis: Visual

• Step Chart

- To visualize a cumulative frequency or **cumulative normalized frequency**.
- Ex. Question – **What proportion of participants did not** always wear the seatbelt?
- Can be used to determine median.

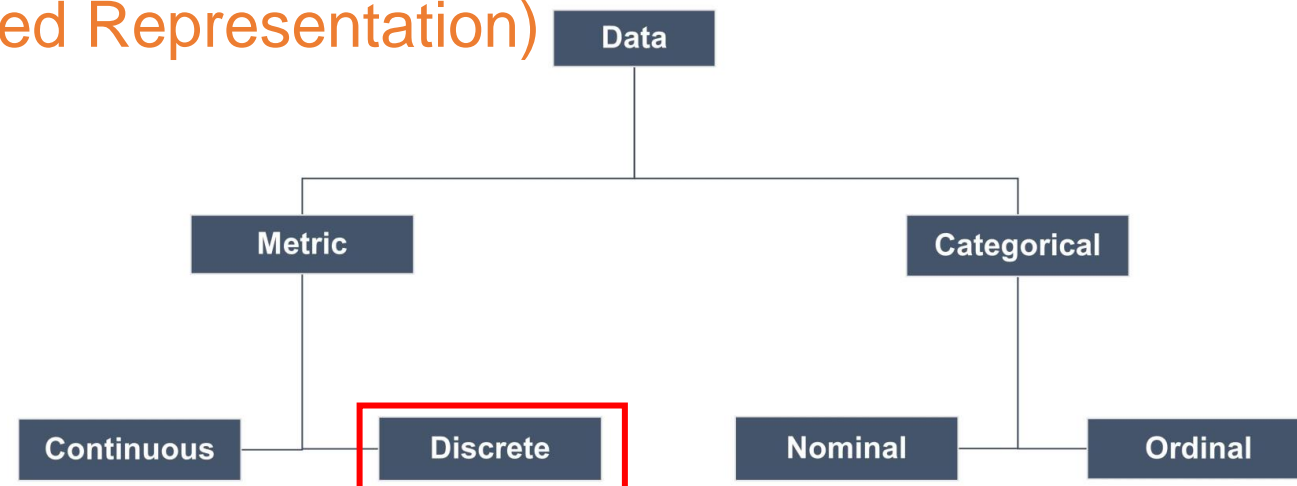
Seatbelt	Frequency	Cumulative Normalized Frequency (%)
Never	0	0
Rarely	2	13.3
Sometimes	1	20.0
Mostimes	5	53.3
Always	7	100



Descriptive Analysis: Visual

- **Discrete Data**

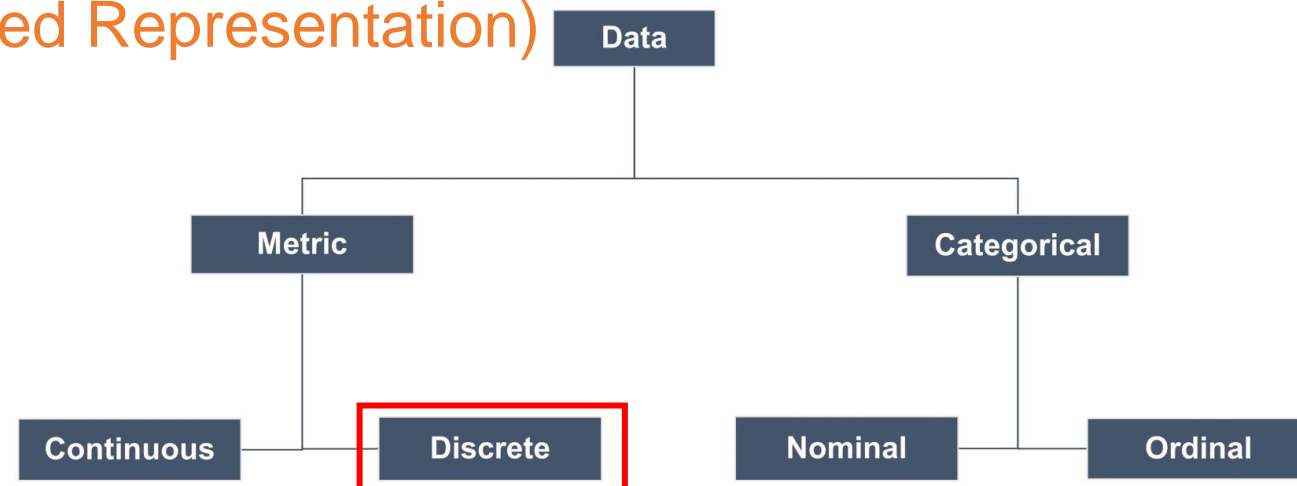
- Pie chart (for a small number of labels) - *Rarely*
- Tag Cloud (for a large number of labels) - *Rarely*
- Bar Chart
- Clustered/Stacked Bar Chart
- Step Chart
- Box-Plot
- Histograms (Grouped Representation)
- Cumulative Histograms (Grouped Representation)



Descriptive Analysis: Visual

- **Continuous Data**

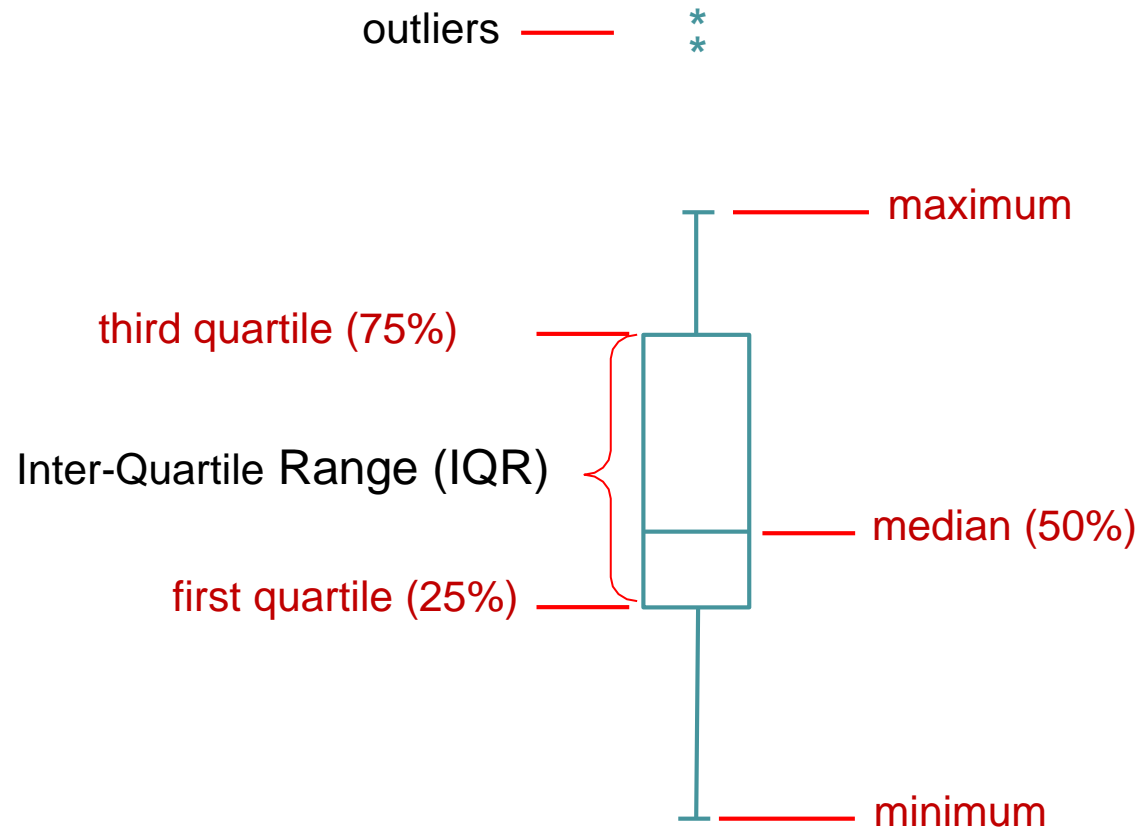
- ~~Pie chart (for a small number of labels)~~
- ~~Tag Cloud (for a large number of labels)~~
- ~~Bar Chart~~
- Clustered/Stacked Bar Chart
- ~~Step Chart~~
- Box-Plot
- Histograms (Grouped Representation)
- Cumulative Histograms (Grouped Representation)



Descriptive Analysis: Visual

- **Box-plot**

- A visual display of the five-number summary.

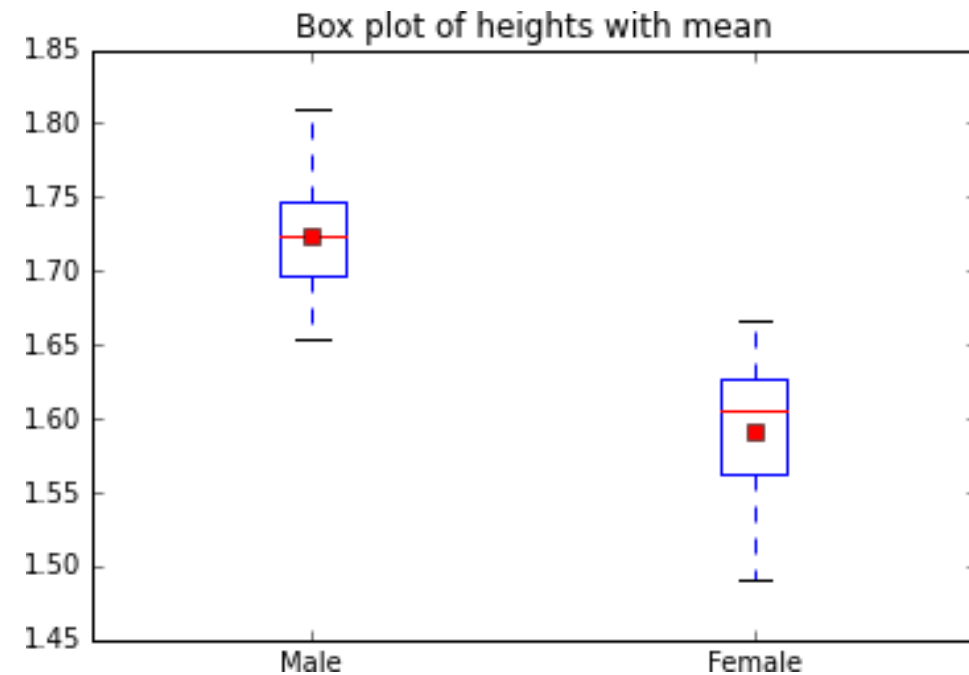


Descriptive Analysis: Visual

- **Box-plot**

- A visual display of the five-number summary.

Participant ID	Gender	Height (m)
1	Male	1.8
2	Female	1.57
3	Male	1.73
4	Male	1.7
5	Male	1.81
6	Female	1.63
7	Male	1.7
8	Female	1.55
9	Female	1.62
10	Female	1.6
11	Male	1.71
12	Male	1.65
13	Female	1.56

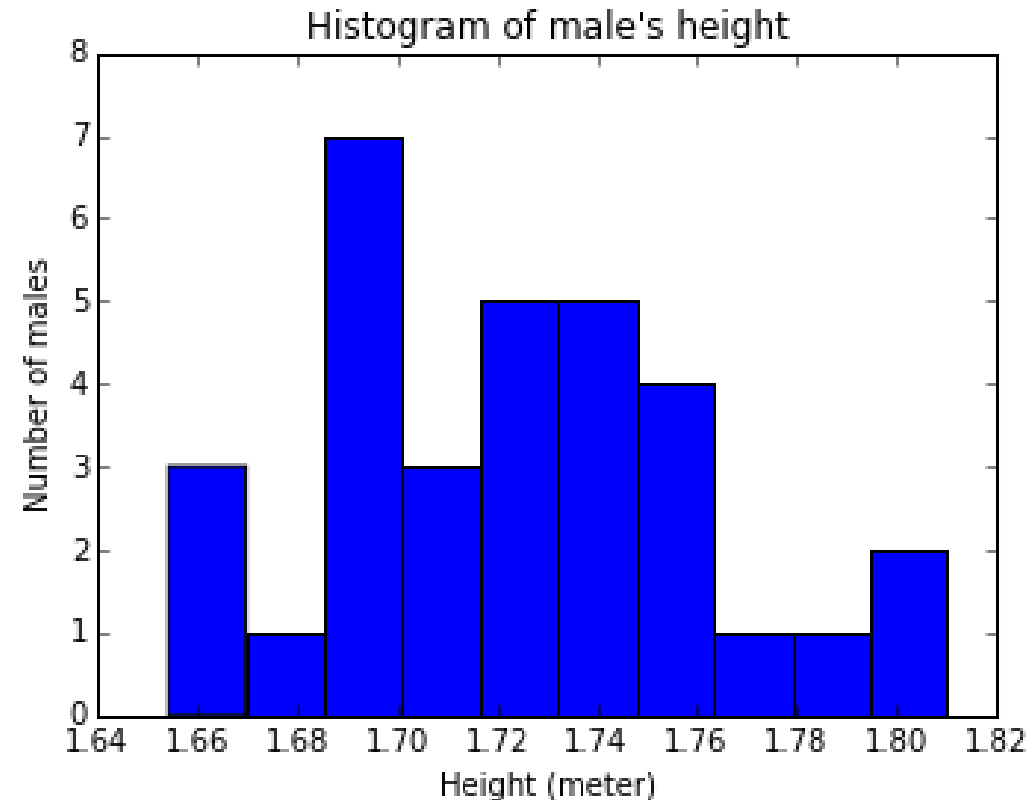


Descriptive Analysis: Visual

- **Histogram**

- Divide data into consecutive bins of equal size.
- Count the number of data points in each bin.
- Plot a bar chart from the resulting data.

Participant	Gender	Hair Color	Height (m)
1	Male	Black	1.797692
2	Female	Red	1.567089
3	Male	Black	1.734498
4	Male	Black	1.696538
5	Male	Black	1.810248
.....
46	Male	Red	1.662167
47	Male	Brown	1.696404
48	Female	Brown	1.597542
49	Male	Black	1.723152
50	Female	Black	1.609744

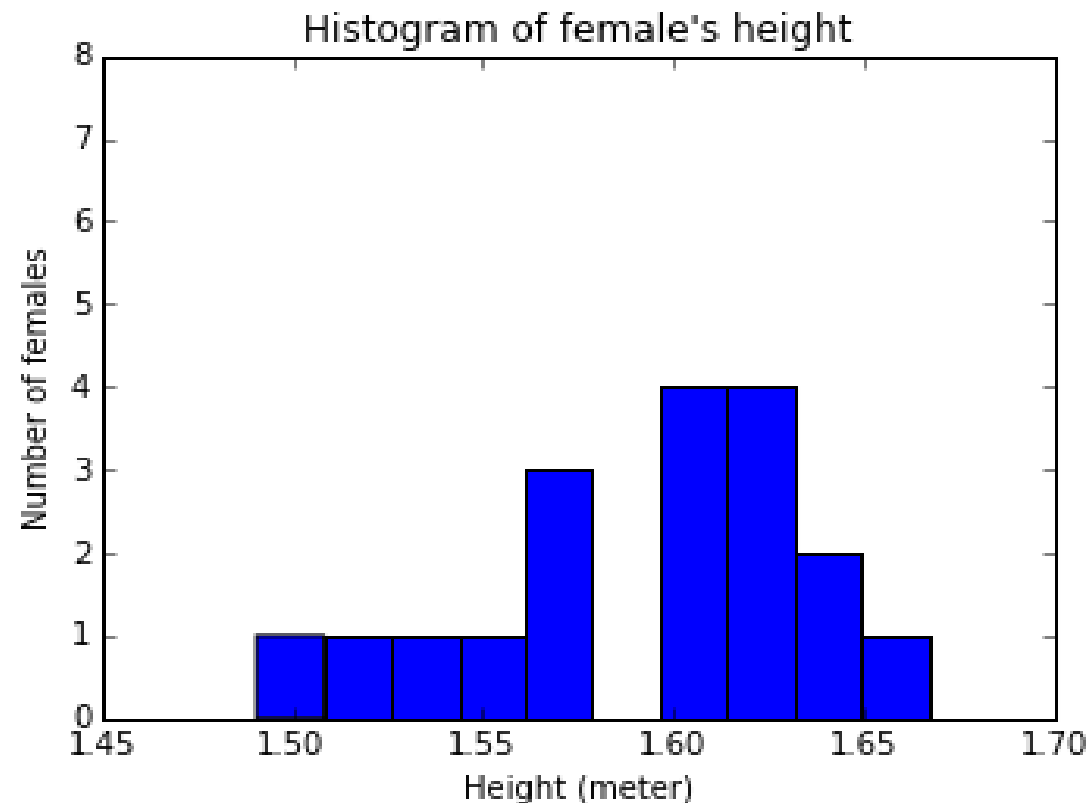


Descriptive Analysis: Visual

- **Histogram**

- Divide data into consecutive bins of equal size.
- Count the number of data points in each bin.
- Plot a bar chart from the resulting data.

Participant	Gender	Hair Color	Height (m)
1	Male	Black	1.797692
2	Female	Red	1.567089
3	Male	Black	1.734498
4	Male	Black	1.696538
5	Male	Black	1.810248
.....
46	Male	Red	1.662167
47	Male	Brown	1.696404
48	Female	Brown	1.597542
49	Male	Black	1.723152
50	Female	Black	1.609744

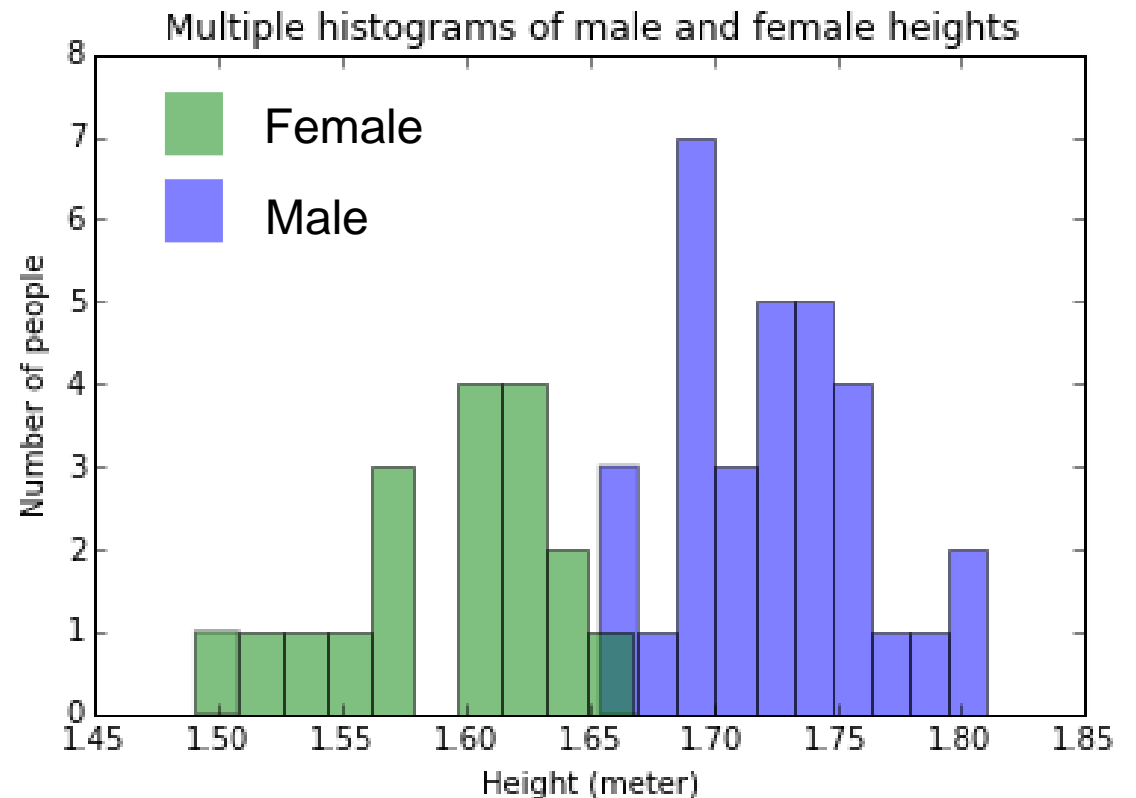


Descriptive Analysis: Visual

- **Multiple Histogram**

- Divide data into consecutive bins of equal size.
- Count the number of data points in each bin.
- Plot a bar chart from the resulting data.

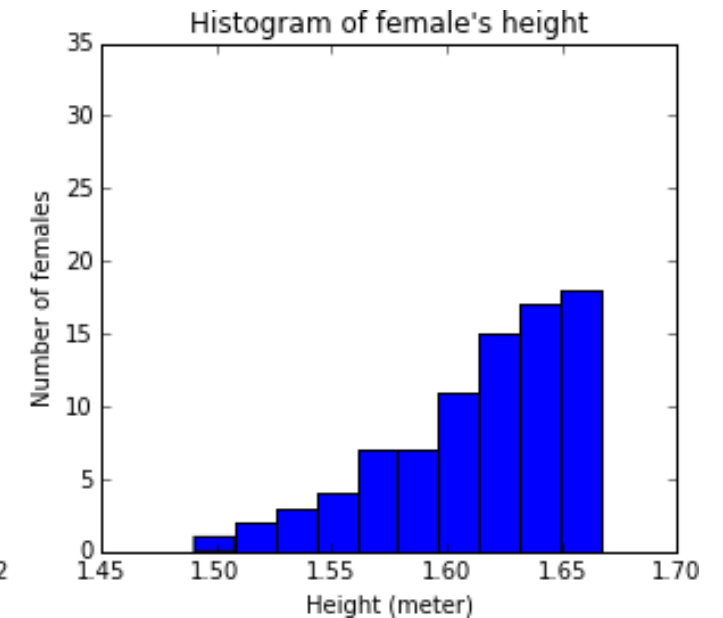
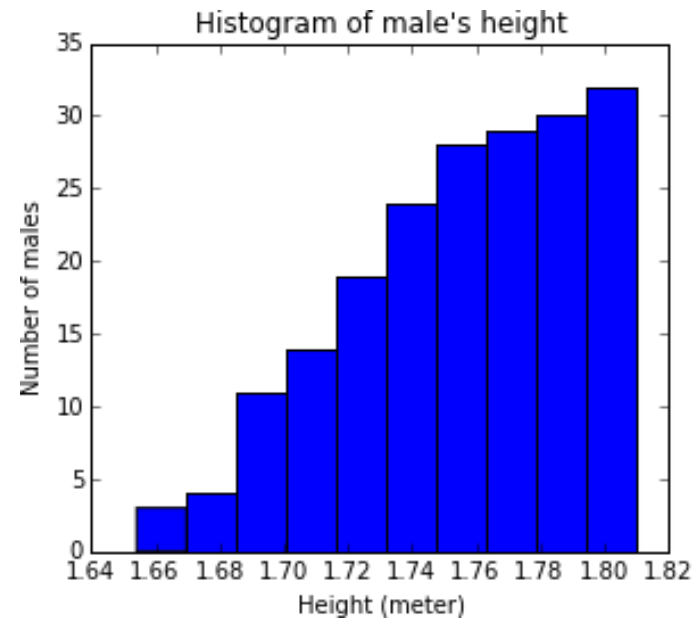
Participant	Gender	Hair Color	Height (m)
1	Male	Black	1.797692
2	Female	Red	1.567089
3	Male	Black	1.734498
4	Male	Black	1.696538
5	Male	Black	1.810248
.....
46	Male	Red	1.662167
47	Male	Brown	1.696404
48	Female	Brown	1.597542
49	Male	Black	1.723152
50	Female	Black	1.609744



Descriptive Analysis: Visual

- Cumulative Histogram
- Cumulative Normalized Histogram
- Cumulative Distribution (=Step Chart for categorical data)

Participant	Gender	Hair Color	Height (m)
1	Male	Black	1.797692
2	Female	Red	1.567089
3	Male	Black	1.734498
4	Male	Black	1.696538
5	Male	Black	1.810248
.....
46	Male	Red	1.662167
47	Male	Brown	1.696404
48	Female	Brown	1.597542
49	Male	Black	1.723152
50	Female	Black	1.609744



Descriptive Analysis: Visual

Attribute Type	Categorical		Metric	
	Nominal	Ordinal	Discrete	Continuous
Pie Chart	Yes	Yes	No	No
Tag Cloud	Yes	Yes	Possible	No
Bar Chart	Yes	Yes	Yes	No
Clustered/Stacked Bar Chart	Yes	Yes	Yes	No
Step Chart	No	Yes	Yes	No
Box Plot	No	No	Yes	Yes
Histogram	No	No	Yes	Yes
Cumulative Histogram	No	No	Yes	Yes

Questions?