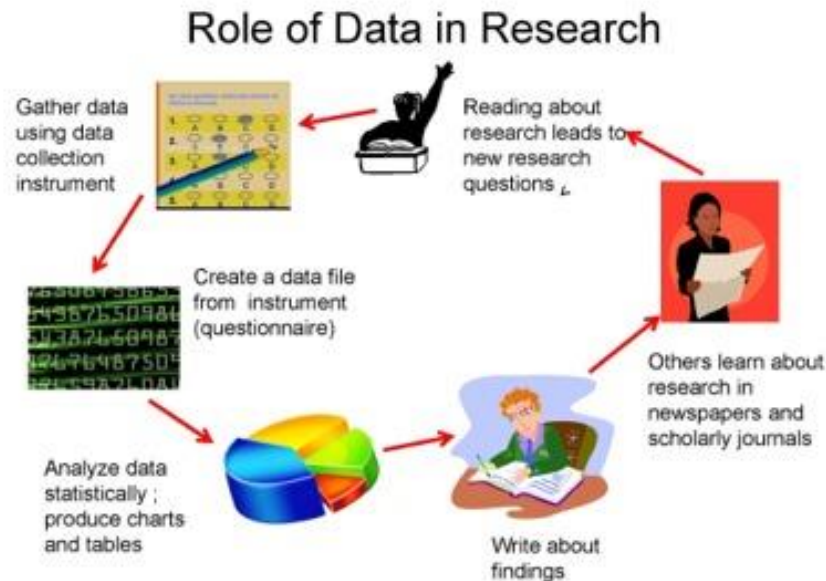# BME 1132
Probability and Biostatistics

Instructor: Ali AJDER, *Ph.D.*

Week-2

- Introduction to Statistics
- Why Statistics are important?
- The role of Statistics in Engineering
- The types of Statistics
- Variables
- Levels of Variables/ Scales
- Summary

Introduction to Statistics

Data/ Information Analysis

Role of Data in Research

"Workers must be equipped not simply with technical know-how, but also with the ability to create, analyze, and transform information and to interact effectively with others. That is, separate the facts from opinions, and then organize these facts in an appropriate manner and analyze the information." **Dr. Alan Greenspan**

'2 out of 3 dentists surveyed indicated they would recommend Brand X toothpaste to their patients.'

"There are two ways of lying. One, not telling the truth and the other, making up statistics." **Josefina Vázquez Mota**

Examples

Governments and organizations,

Life expectancy of a population,
The risk factors for infant mortality,
How a political system affects trade policy,
Who is going to vote for a political party in the next election,
Geographical differences in energy usage, migration patterns, or reasons for unemployment….

In business,

Identifying people who may be interested in a certain product,
Optimizing prices, and
Evaluating the satisfaction of customers are possible areas of interest…

In BIOSTATISTICS,

Using the tools of statistics, biostatisticians help answer pressing research questions in medicine, biology and public health, such as
whether a medical intervention helps in reducing the burden of a disease,
what causes cancer and other diseases,
how long a person with a certain illness is likely to survive,
how personality relates to decision-making,
whether a new fertilizer increases the yield of crops,
what are the long-term changes in the population of a fish species…

Why study Statistics?

There are at least three reasons for studying statistics…

1) Data are everywhere,

2) Statistical techniques are used to make many decisions that affect our lives, and

3) No matter what your career, you will make professional decisions that involve data.

An understanding of statistical methods will help you make these decisions more effectively.

We spend our day producing Data…

10 minutes

"We create as much information in two days now as we did from the dawn of man through 2003"

(2003) Eric Schmidt, former CEO at Google

2015

STARTUP SATURDAY
progettare·innovare·condividere

We spend our day producing Data…

Eric Schmidt's "5 Exabytes" Quote is a Load of Crap

Robert J. Moore
February 7, 2011

[follow us on Twitter at **@RJMetrics** and **visit our website**
the true power of your data]

Let me start out by saying that I'm a huge fan of Google's
however, I was a little perplexed by a statement he made
convention. He said the following:

"There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days."
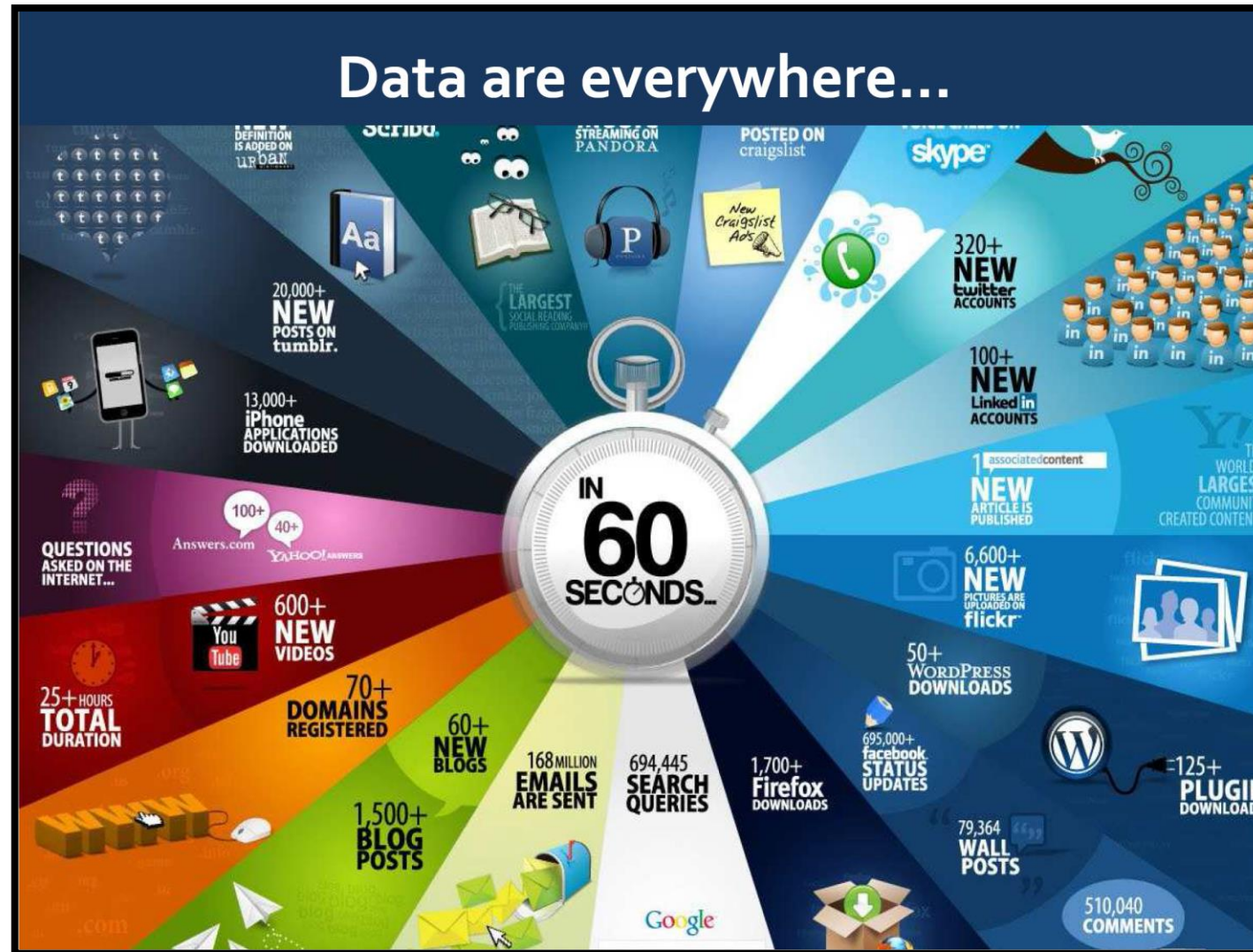

From Wikipedia, the free encyclopedia

The **exabyte** is a multiple of the unit byte for digital information. In the International System of Units (SI), the prefix *exa* indicates multiplication by the sixth power of 1000 (10^{18}). Therefore, one exabyte is one quintillion bytes (short scale). The symbol for the exabyte is **EB**.

1 EB = 1000^6 bytes = 10^{18} bytes = 1 000 000 000 000 000 000 B = 1000 petabytes = 1 million terabytes = 1 billion gigabytes.
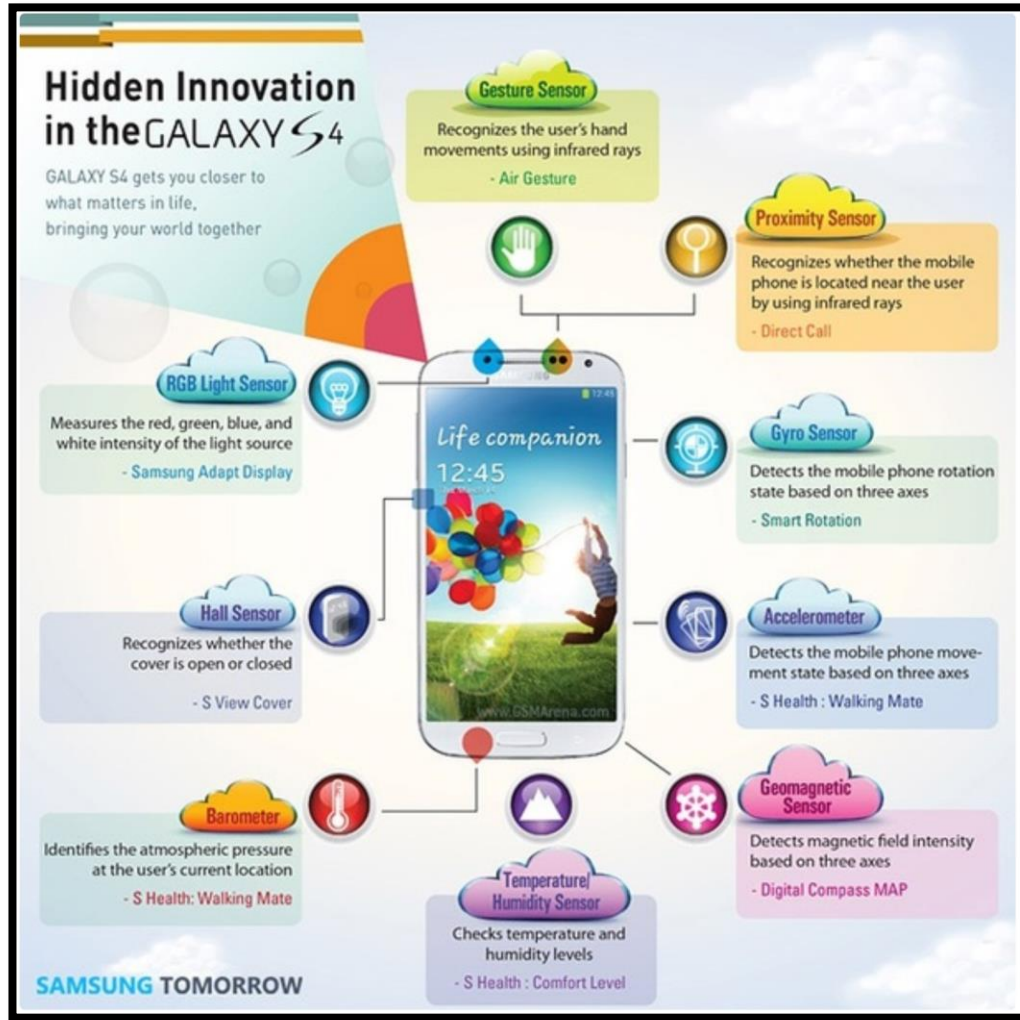
A related unit, the exbibyte, using a binary prefix, is equal to 1024^6 (=2^{60}) bytes, about 15% larger.

One thousand exabytes (1000 EB) is equal to one zettabyte (1 ZB).

"23 Exabytes of information was recorded and replicated in 2002. We now record and transfer that much information every 7 days."
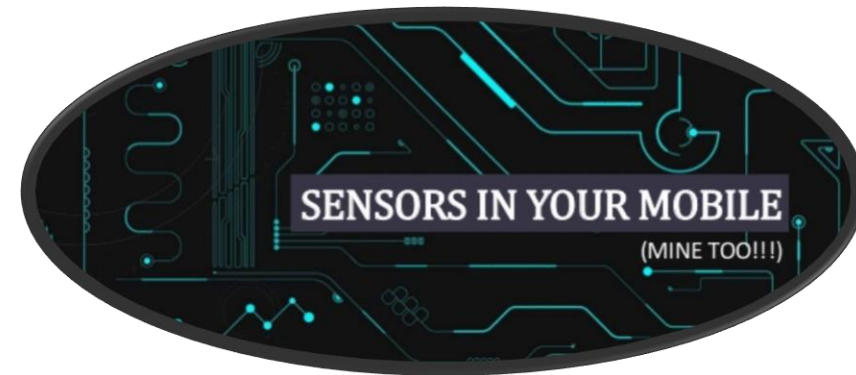
Data are everywhere…
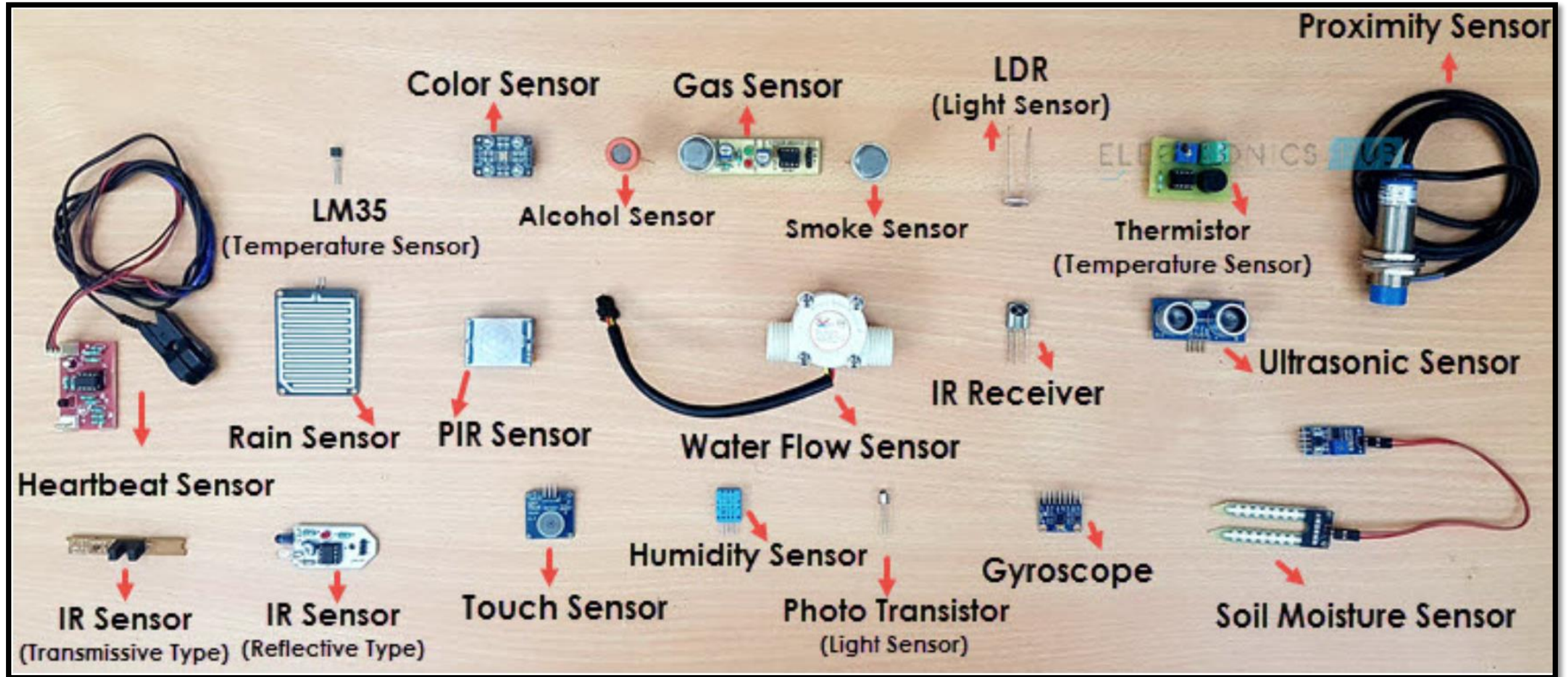
Data are everywhere…

Samsung Galaxy Note 10+
Accelerometer
Barometer
Ultrasonic fingerprint scanner
Pressure sensor
Magnetometer
Gyroscope
Hall sensor
Proximity sensor
RGB light sensor

Iphone 11 Pro & Iphone 11 Pro Max
Face ID
Barometer
Three-axis gyro
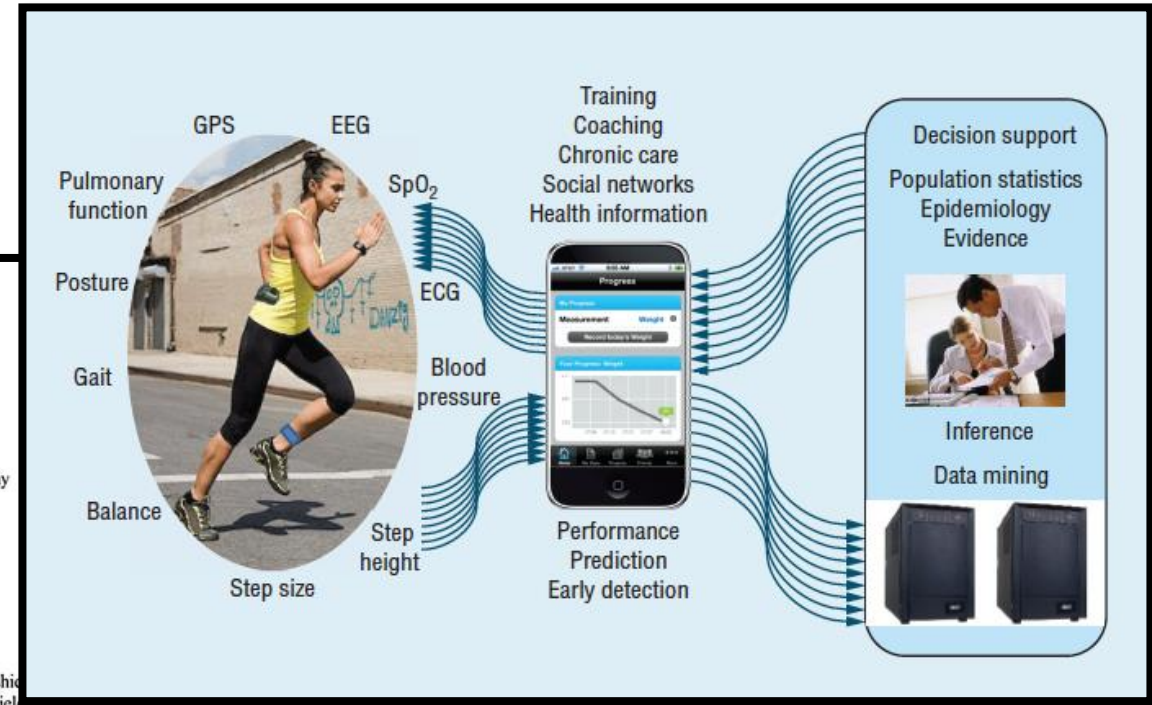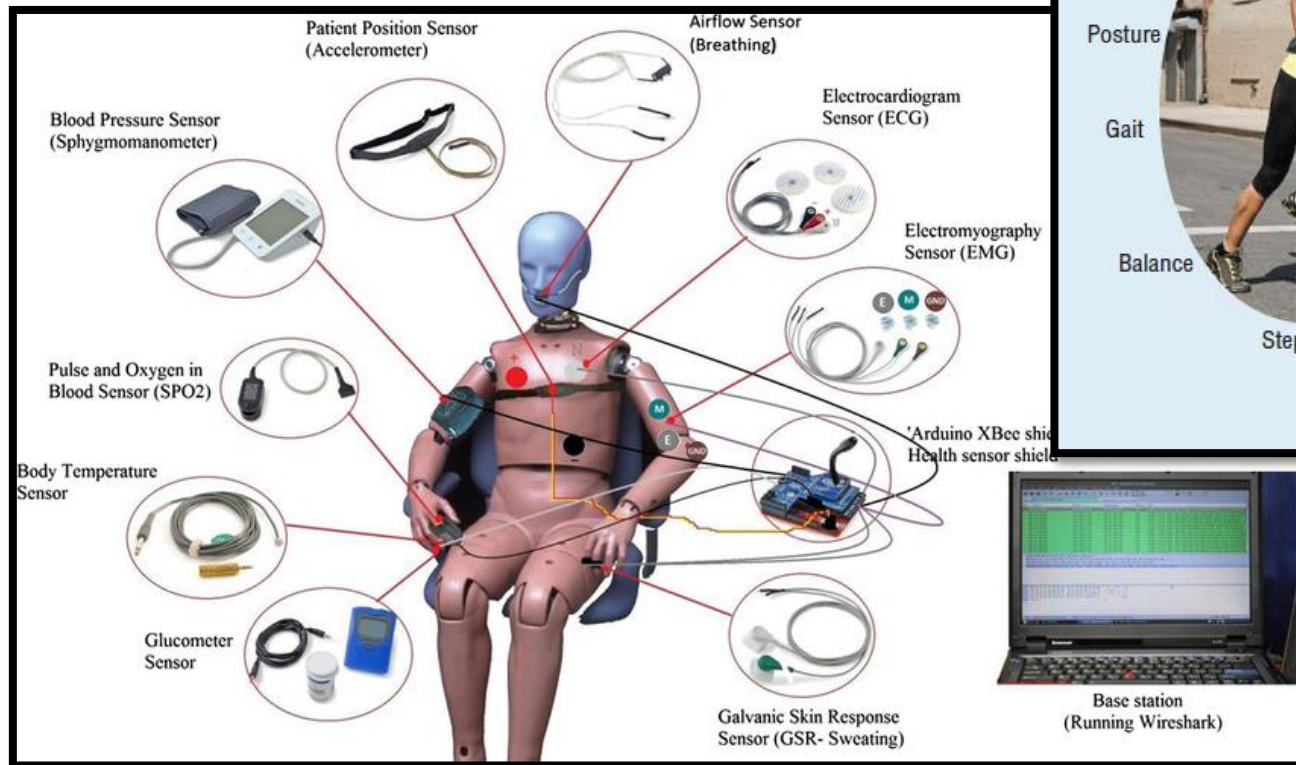Accelerometer
Proximity sensor
Ambient light sensor
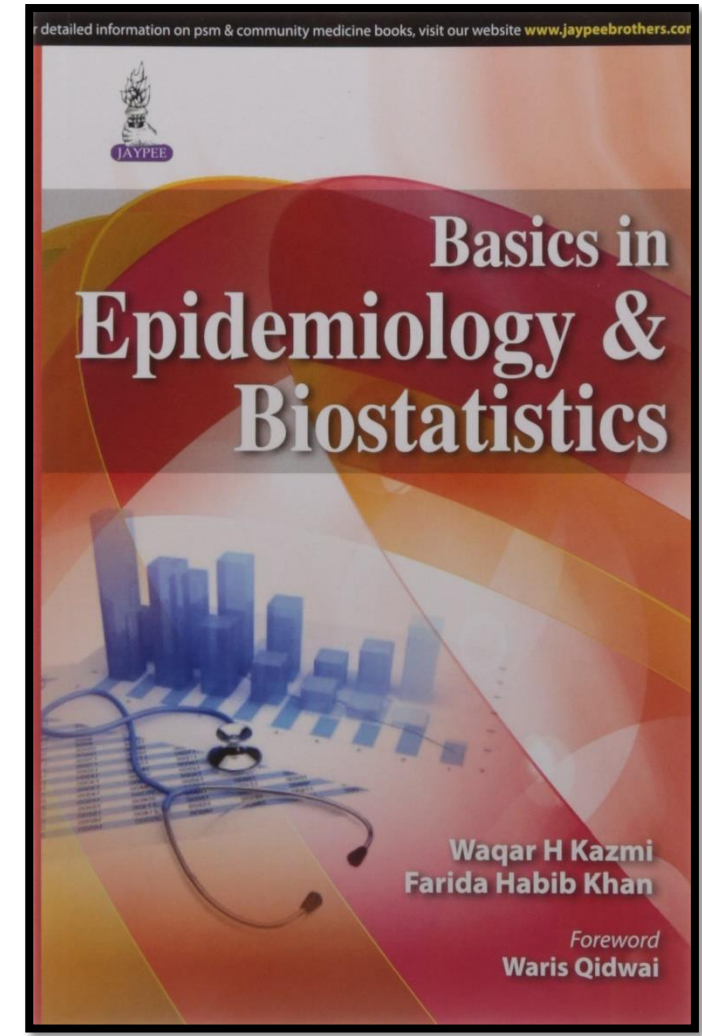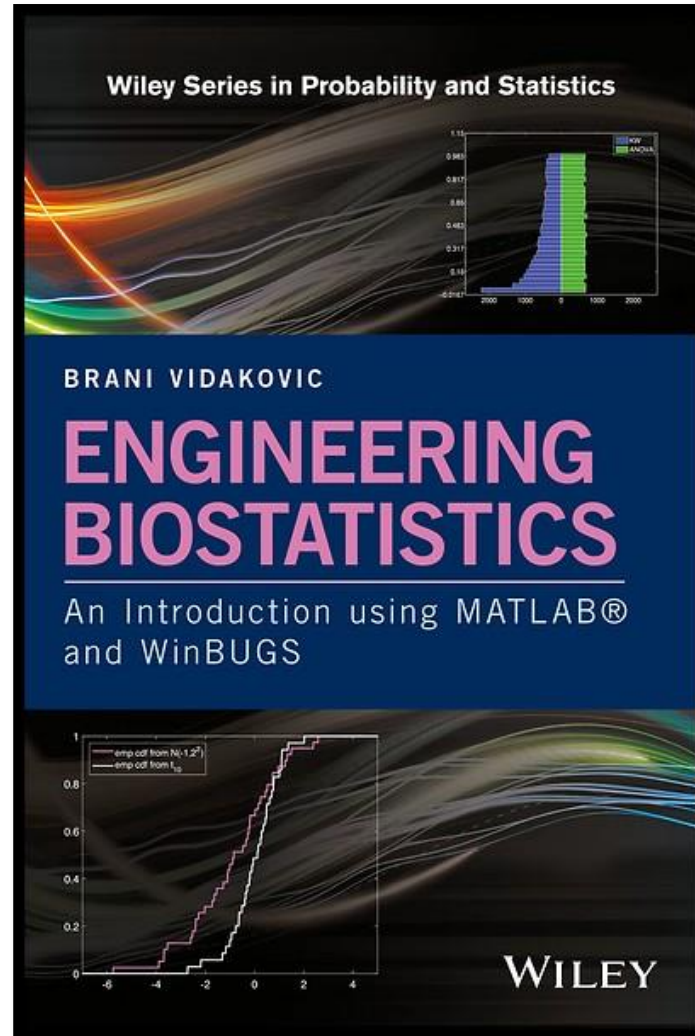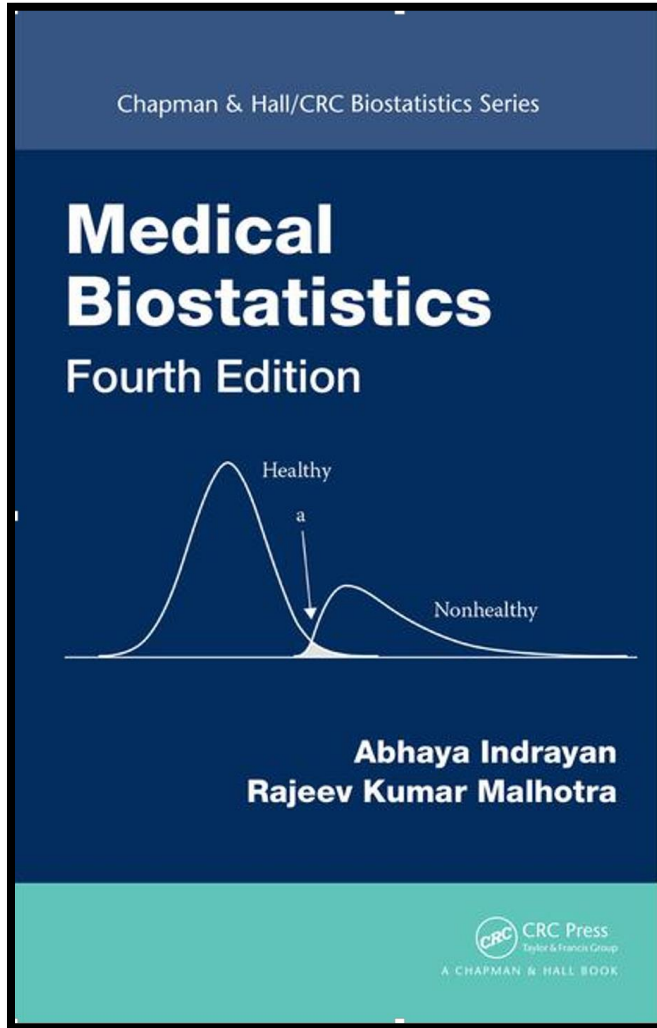
Data are everywhere… Engineering…

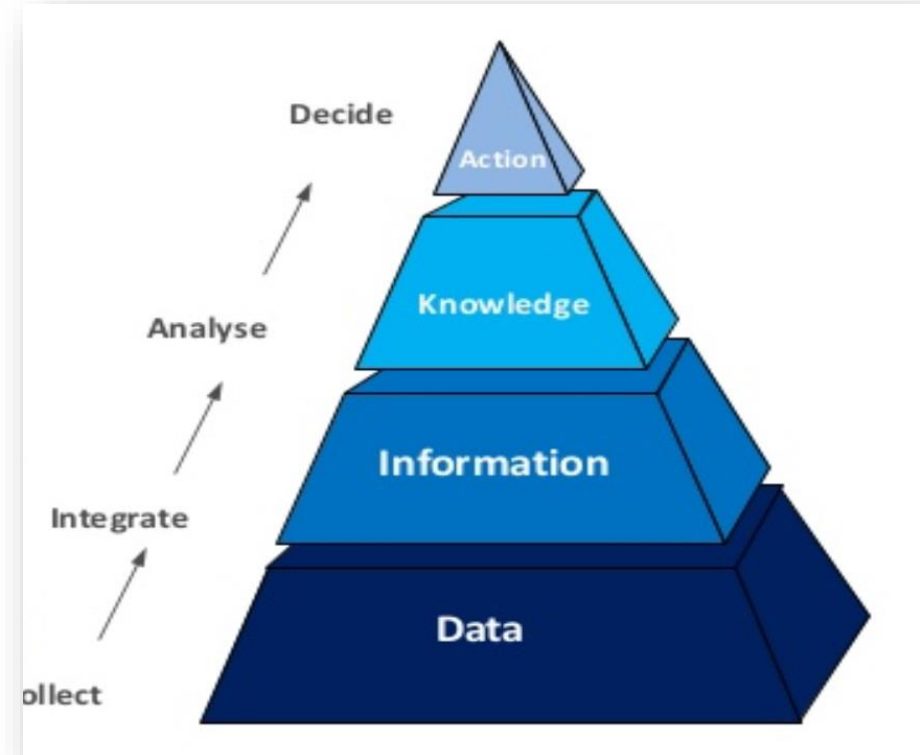# Data are everywhere… Biostatistics…

E-Health Sensors

Evidence-Based Practice (EBP)

The Role of Statistics in Biomedical Engineering

Decision making from data…
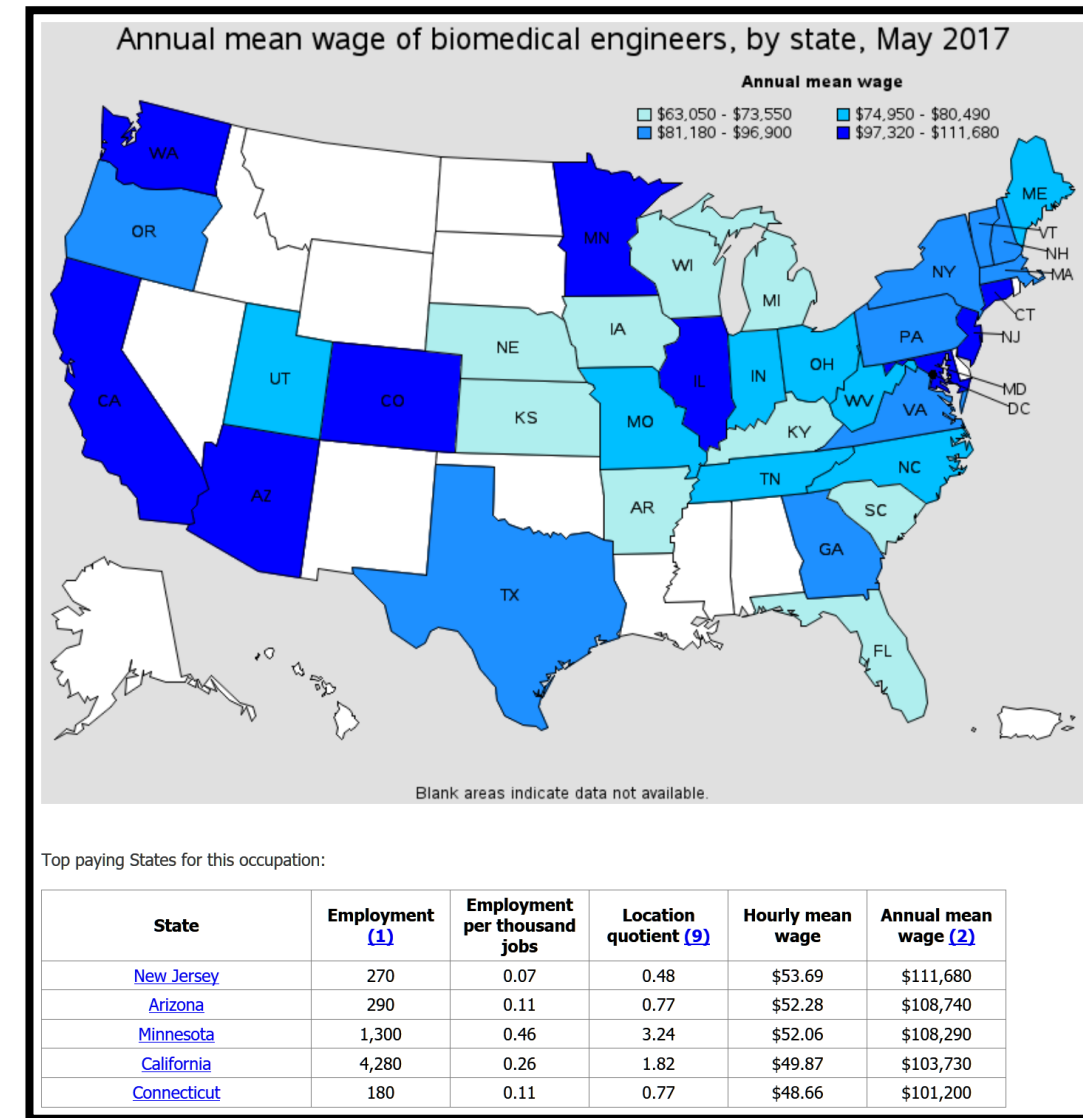
What is meant by Statistics?

It really has two meanings.

1. In the more common usage, statistics refers to **numerical information.**

2. In this lecture, it has a much broader meaning than just collecting and publishing numerical information;

DEFINITION OF STATISTICS

Annual mean wage of biomedical engineers, by state, May 2017

Top paying States for this occupation:

State	Employment (1)	Employment per thousand jobs	Location quotient (9)	Hourly mean wage	Annual mean wage (2)
New Jersey	270	0.07	0.48	$53.69	$111,680
Arizona	290	0.11	0.77	$52.28	$108,740
Minnesota	1,300	0.46	3.24	$52.06	$108,290
California	4,280	0.26	1.82	$49.87	$103,730
Connecticut	180	0.11	0.77	$48.66	$101,200

Types of Statistics
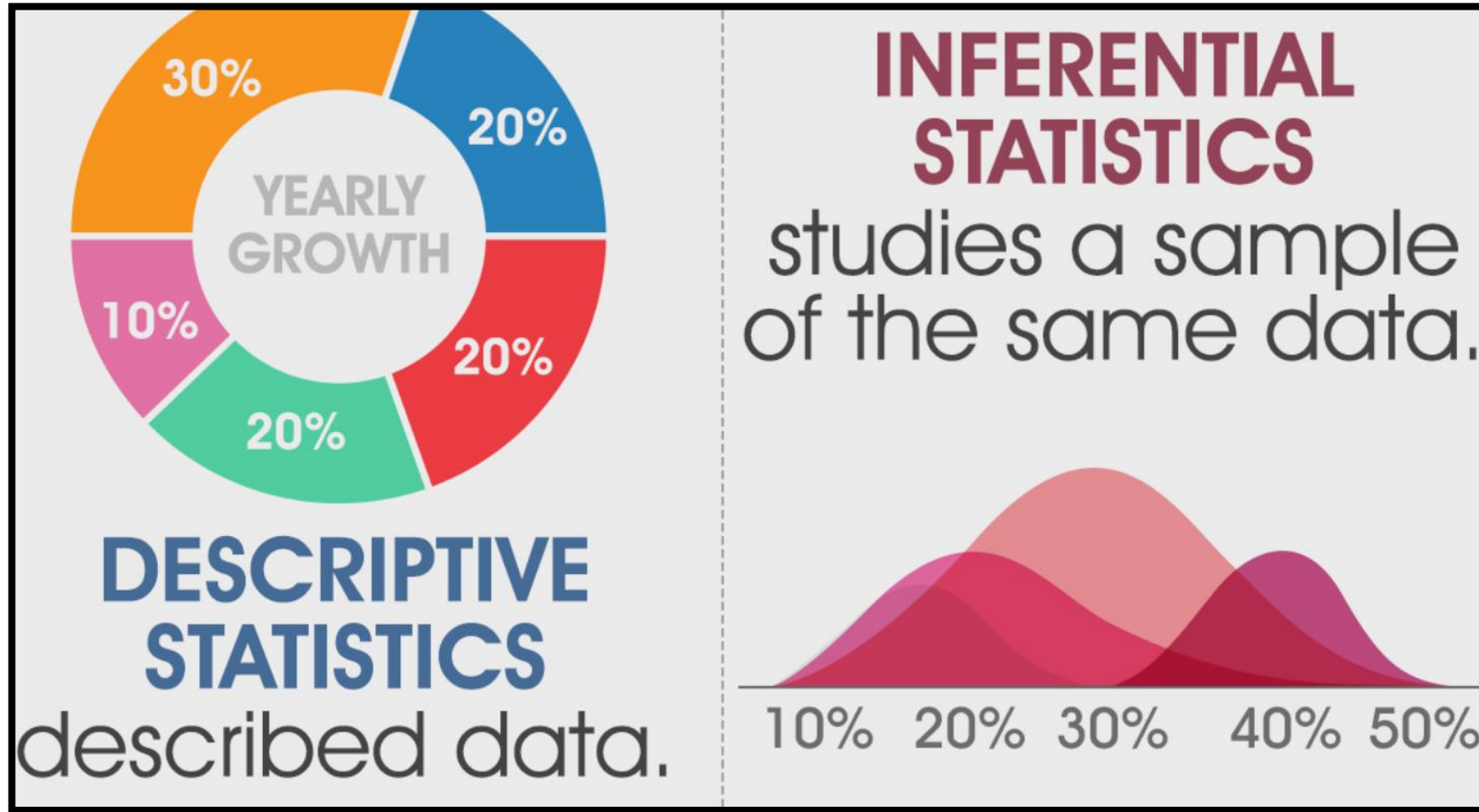
Basic Definitions

Parameter
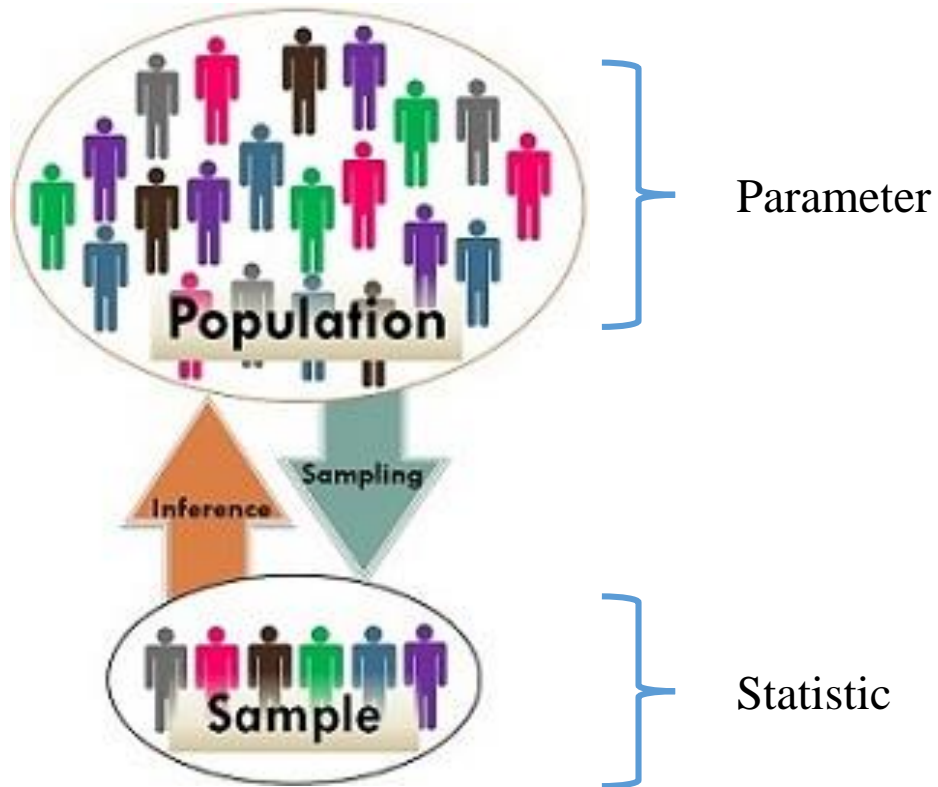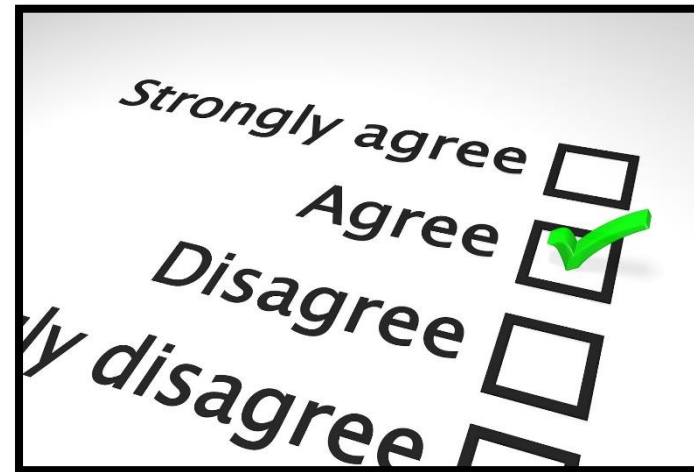
Statistic

- **Population-** $\omega_1, \omega_2, \dots \omega_n \subseteq \Omega$

Population of Turkey?

- **Sample-** $\omega_1, \omega_2, \dots \omega_n$

Why take a *sample* instead of studying every member of the *population*?

Data Collection Methods

➢ Survey

➢ Experiment

➢ Observational Data

➢ Primary and Secondary Data

Data Collection in Engineering

➢ A **retrospective** study using historical data

➢ An **observational** study

➢ A **designed experiment**

Statistical inference is one type of reasoning

Physical laws

Product designs

Types of reasoning

Population

Sample

Statistical inference
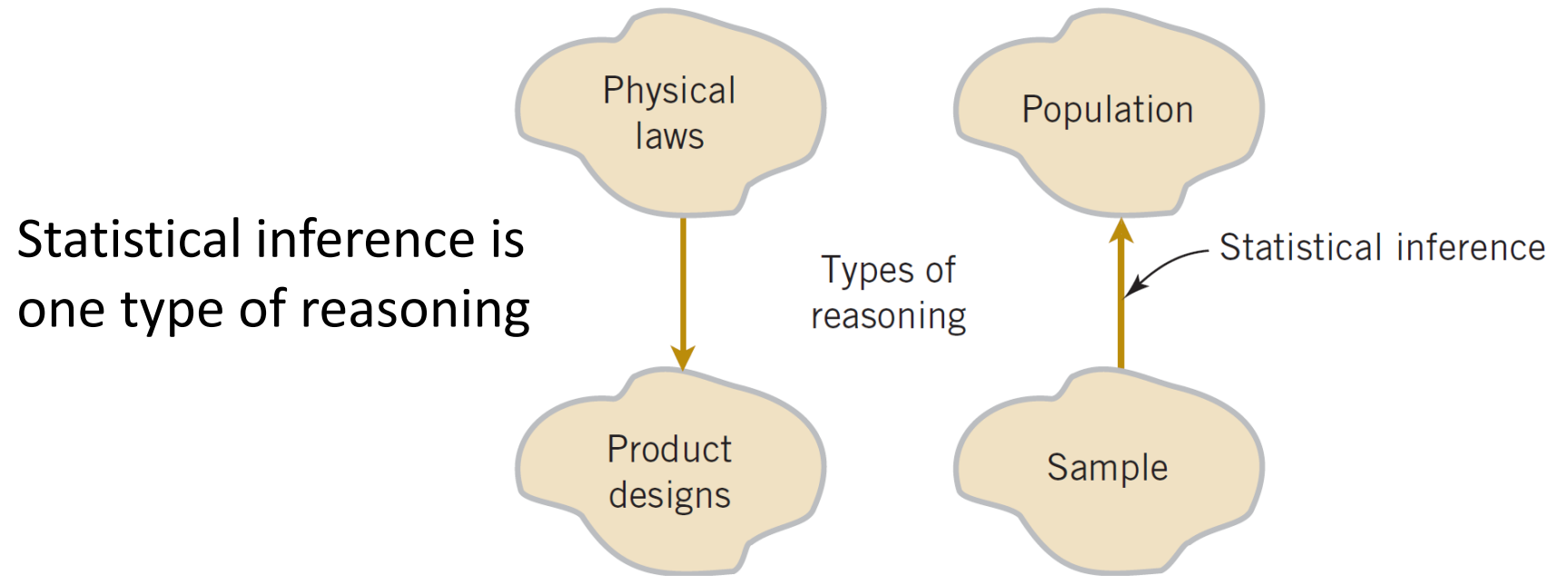
Variable, Observation and Data Set

Types of Variables
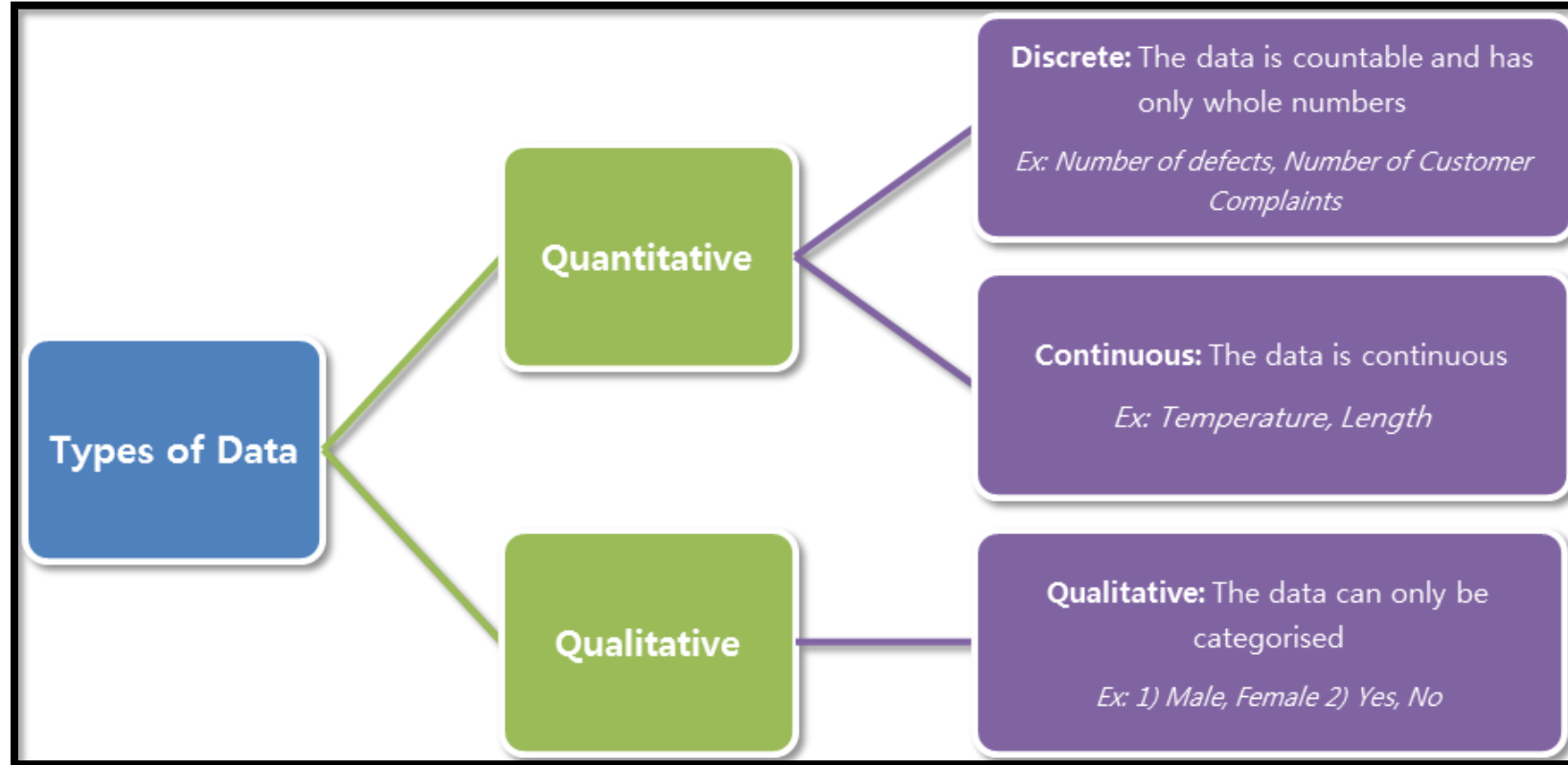
Levels of Variables

1. Nominal
- can only be classified
- can **not** be ordered
- (Colours of M&M chocolate)

2. Ordinal
- can be ordered
- (Rate of instructor in this class/ Good-Average-Poor)

3. Interval
- all the characteristics of the ordinal level
- between values is a constant
- (temperature)

4. Ratio
- all the characteristics of the interval level
- 0 point and ratio of 2 numbers is meaningful
- (price)

Summary-1

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

There are two types of statistics.
A. Descriptive statistics are procedures used to organize and summarize data.
B. Inferential statistics involve taking a sample from a population and making estimates about a population based on the sample results.
1. A population is an entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.
2. A sample is a part of the population.

There are two types of variables.
A. A qualitative variable is nonnumeric.
1. Usually we are interested in the number or percent of the observations in each category.
2. Qualitative data are usually summarized in graphs and bar charts.
B. There are two types of quantitative variables and they are usually reported numerically.
1. Discrete variables can assume only certain values, and there are usually gaps between values.
2. A continuous variable can assume any value within a specified range.

Summary-2

The definition of Variability and Uncertainty.
Variability is the extent to which data points in a statistical distribution or data set diverge from the average, or mean, value as well as the extent to which these data points differ from each other.
Uncertainty is a state of doubt about the future or about what is the right thing to do.

There are four levels of measurement.
A. With the nominal level, the data are sorted into categories with no particular order to the categories.
1. The categories are mutually exclusive. An individual or object appears in only one category.
2. The categories are exhaustive. An individual or object appears in at least one of the categories.
B. The ordinal level of measurement presumes that one classification is ranked higher than another.
C. The interval level of measurement has the ranking characteristic of the ordinal level of measurement plus the characteristic that the distance between values is a constant size.
D. The ratio level of measurement has all the characteristics of the interval level, plus there is a zero point and the ratio of two values is meaningful.

Questions?