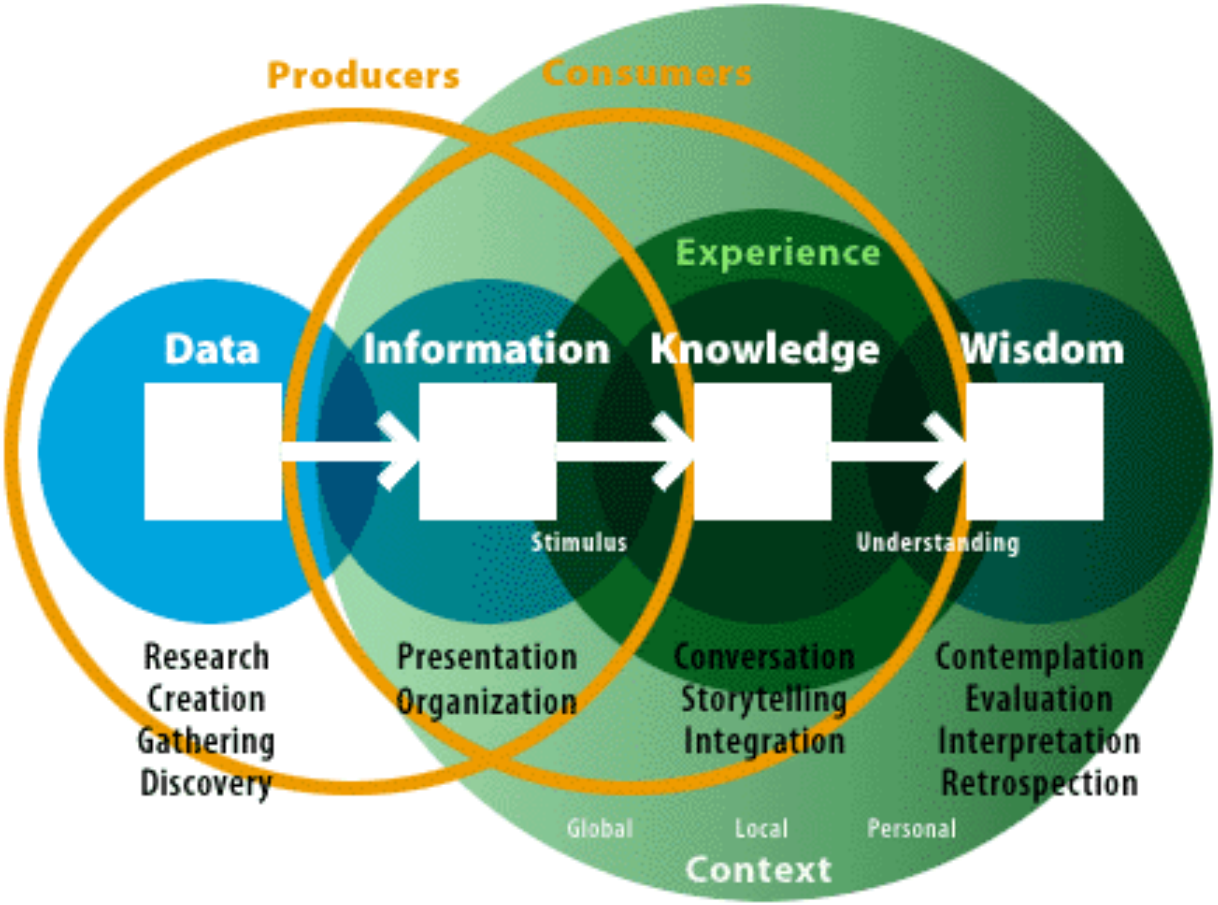
Data - Intelligence

ID 413: Information Graphics and Data Visualization
Spring 2016

Venkatesh Rajamanickam (@venkatrajam)
venkatra@iitb.ac.in
http://info-design-lab.github.io/ID413-DataViz/

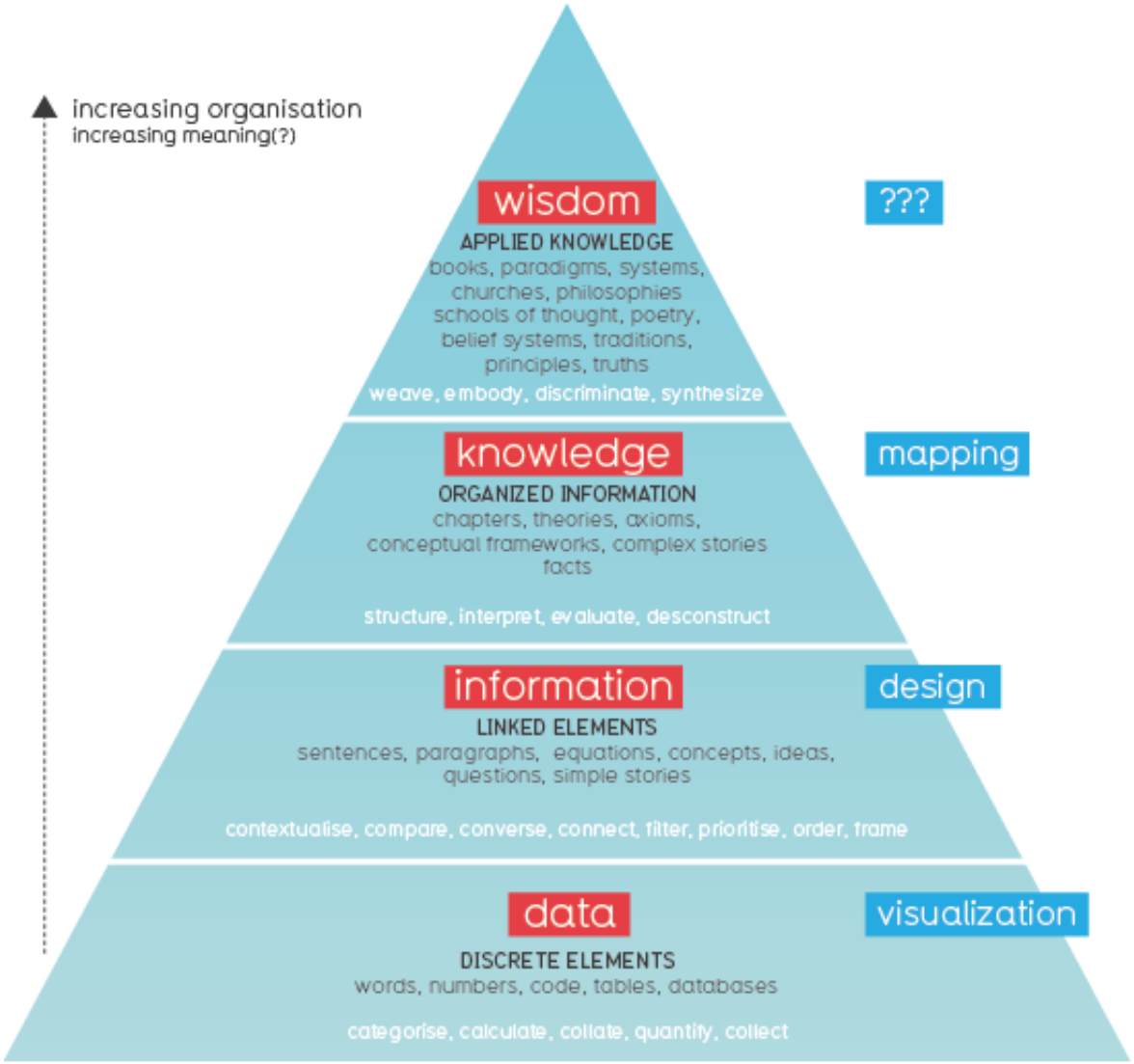
Intelligence Hierarchy: Data, Information, Knowledge, Wisdom



Nathan Shedroff

Hierarchy Of Visual Understanding?

Just playing. Something in this?



David McCandless // v 0.1 // work in progress
InformationIsBeautiful.net

Measurement is the foundation of all data

"Whenever you can, count."
— Sir Francis Bacon

"One accurate measurement is worth a thousand expert opinions."
— Grace Hopper

"In many spheres of human endeavour, from science to business to education to economic policy, good decisions depend on good measurement."
— Ben Bernanke

"I have been struck again and again by how important measurement is to improving the human condition."
— Bill Gates

Measurement is the foundation of all data

"Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it."
— H. James Harrington

"If a measurement matters at all, it is because it must have some conceivable effect on decisions and behaviour. If we can't identify a decision that could be affected by a proposed measurement and how it could change those decisions, then the measurement simply has no value"
— Douglas W. Hubbard, *How to Measure Anything: Finding the Value of Intangibles in Business*

How Measurement led to the Modern World

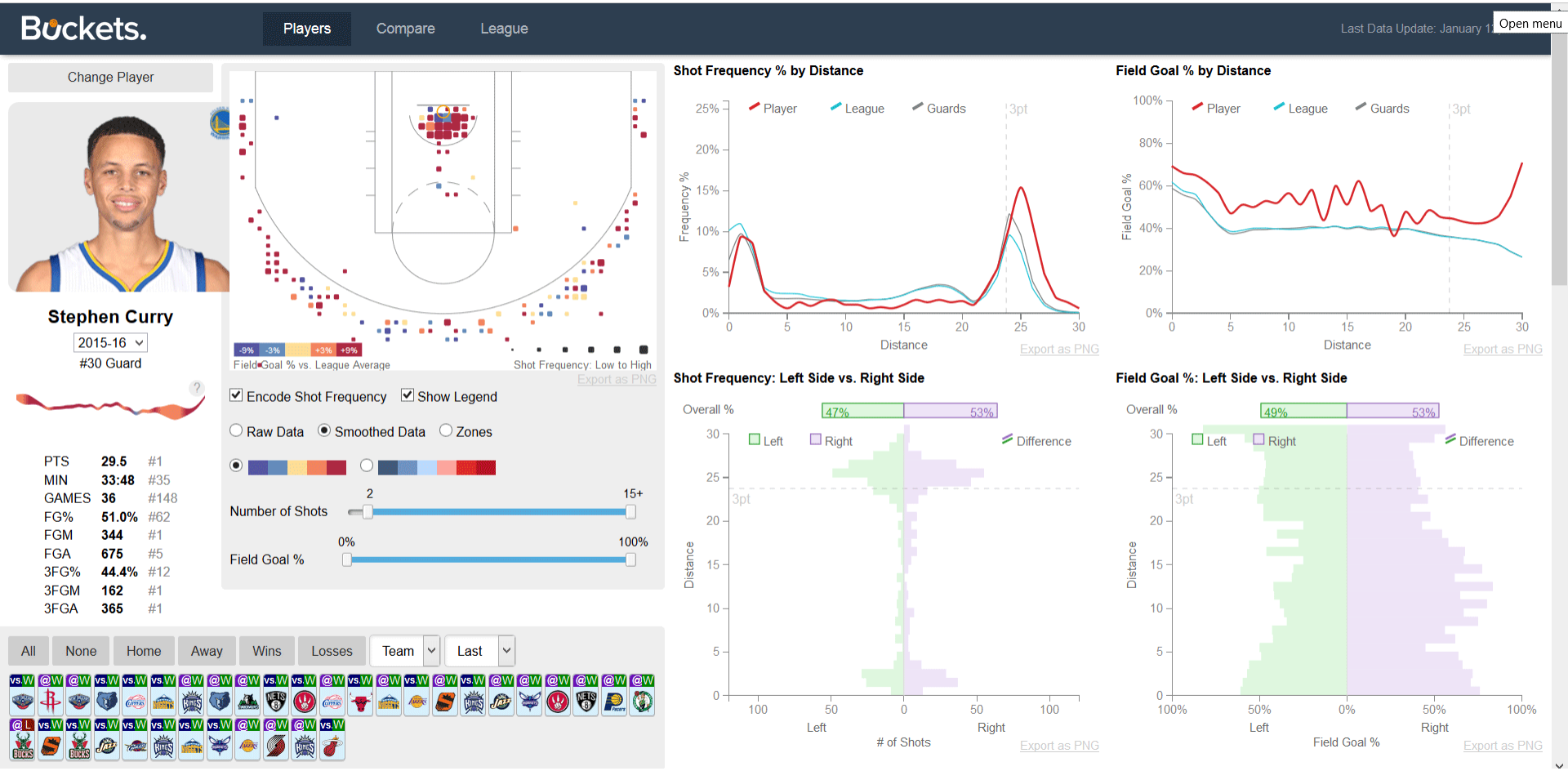
Pre-modern customs were all about dealing with trust, the need for direct supervision, and facing up to the enormous risks posed by nature.

Once fundamental measurement problems were solved — involving time, distance, weights, and power, among others — it became possible to cheaply measure worker performance, input and output quality, and the role of nature, in areas of life that were unheard of before.

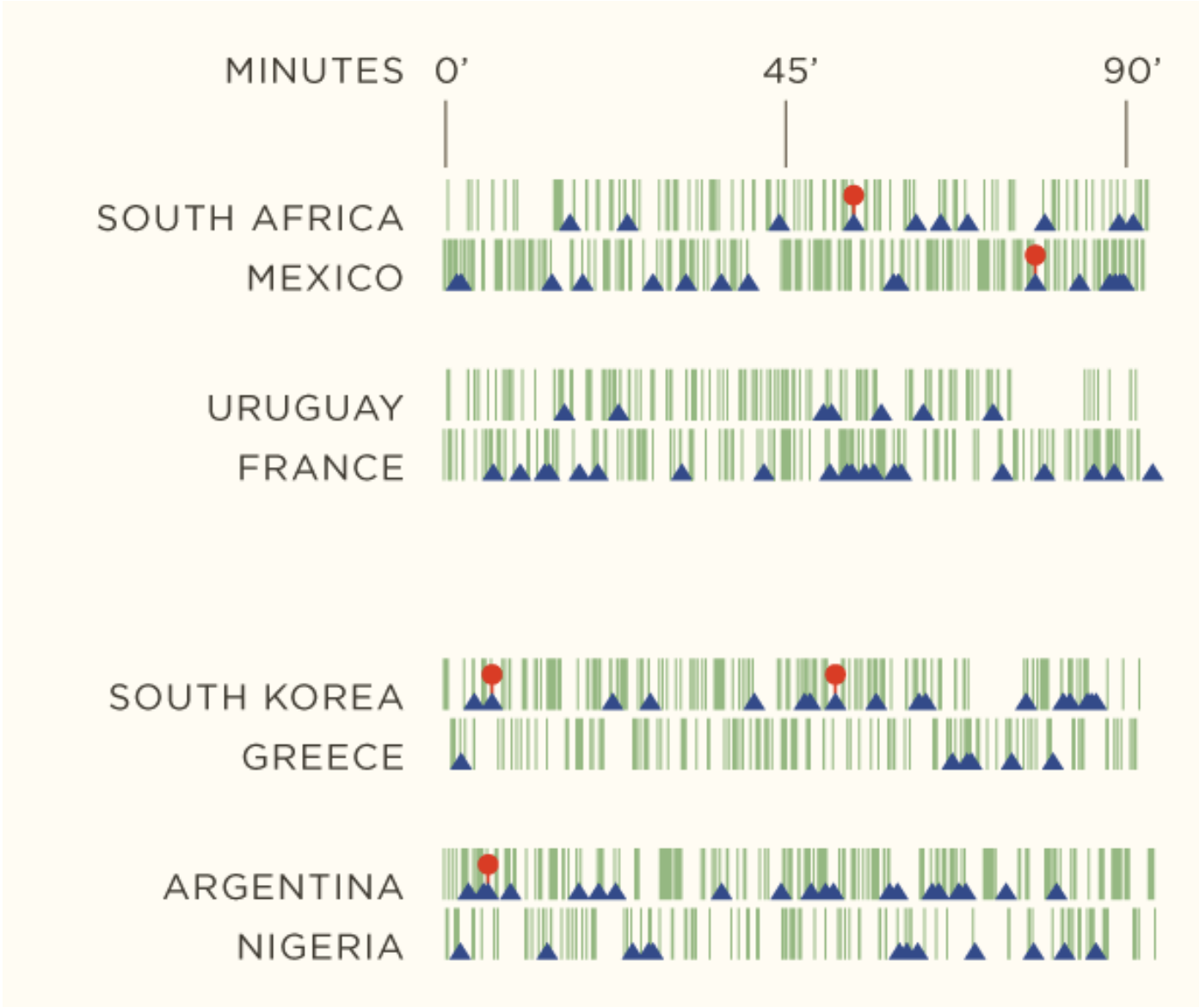
This ability to cheaply measure ushered in the world of modern institutions.

— The Institutional Revolution: Measurement and the Economic Emergence of the Modern World (2011) by Douglas W. Allen

Why is it hard to measure the value of soccer players?



Why is it hard to measure the value of soccer players?



Why is it hard to measure the value of soccer players?

- So what is it about soccer that makes it so hard to quantify?
- Or what makes cricket so easy to measure?

Why is it hard to measure the value of soccer players?

- So what is it about soccer that makes it so hard to quantify?
- Or what makes cricket so easy to measure?
- One obvious answer is the quantity of the units that can be easily separated and analysed.

Why is it hard to measure the value of soccer players?

- So what is it about soccer that makes it so hard to quantify?
- Or what makes cricket so easy to measure?
- One obvious answer is the quantity of the units that can be easily separated and analysed.
- In cricket it is innings, overs, balls, runs, wickets, catches and so on; in basketball it is points, assists, rebounds, steals, turnovers, etc., in tennis it is serves, volleys, aces, errors (forced, unforced) etc.
- For soccer, the only apparent unit to separate out is the 45 minute halftime mark. Changes in possession could be another measure, or number of passes, or shots at goal etc., but none are very useful.

(Lionel Messi: http://fivethirtyeight.com/features/lionel-messi-is-impossible/
Stephen Currry: http://peterbeshai.com/buckets/app/#/playerView/201939_2015)

Measurement

SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues

by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that

On the Theory of Scales of Measurement. S. S. Stevens. Science, 103 (2684), pp. 677-680, June 1946.

Level of measurement

Level of measurement or scale of measure is a classification that describes the nature of information within the numbers assigned to variables.

The best known classification is that developed by the psychologist **Stanley Smith Stevens**, who proposed four levels, or scales, of measurement: **nominal, ordinal, interval,** and **ratio.**

Stevens proposed his typology in a 1946 Science article titled "*On the theory of scales of measurement*", in which he claimed that all measurement can be conducted using the four different types of scales, unifying both **qualitative** and **quantitative** data.

Nominal Scale

Nominal scales are used for labeling variables, without any quantitative value.

“Nominal” scales could simply be called “labels.”

Examples include, gender, nationality, ethnicity, language, genre, style, biological species, and form.

Notice that all of these scales are mutually exclusive (no overlap) and none of them have any numerical significance.

Ordinal Scale

With ordinal scales, it is the order of the values is what's important and significant, but the differences between each one is not really known.

Examples include, values such as 'sick' vs. 'healthy' when measuring health, to a spectrum of values, such as Very Happy', Happy', OK', 'Unhappy' when measuring satisfaction.

Here the scale allows for rank order (1st, 2nd, 3rd, etc.) by which data can be sorted, but still does not allow for relative degree of difference between them.

For example, is the difference between "OK" and "Unhappy" the same as the difference between "Very Happy" and "Happy?" We can't say.

Ordinal scales are typically measures of non-numeric concepts like satisfaction, happiness, discomfort, etc.

Interval Scale

Interval scales are numeric scales in which we know not only the order, but also the exact differences between the values.

Examples include temperature with the Celsius scale, which has two defined points, separated into 100 intervals, dates, percentages, location in Cartesian coordinates, and direction measured in degrees from true or magnetic north.

Ratios are not allowed since 20 degree C cannot be said to be "twice as hot" as 10 degree C, nor can multiplication/division be carried out between any two dates directly.

However, ratios of differences can be expressed; for example, one difference can be twice another.

The realm of statistical analysis on these data sets opens up because central tendency can be measured by mode, median, or mean; and standard deviation can also be calculated.

Interval Scale

Interval scales are numeric scales in which we know not only the order, but also the exact differences between the values.

Examples include temperature with the Celsius scale, time, dates, percentages, location in Cartesian coordinates, and direction measured in degrees from true or magnetic north.

Interval scales don’t have a true zero, and so it is impossible to compute ratios. 20 degree C cannot be said to be "twice as hot" as 10 degree C, nor can multiplication/division be carried out between any two dates directly.

Interval Scale

However, ratios of differences can be expressed; for example, one difference can be twice another.

The realm of statistical analysis on these data sets opens up because central tendency can be measured by mode, median, or mean; and standard deviation can also be calculated.

Ratio Scale

Ratio scales are the ultimate nirvana when it comes to measurement scales because they tell us about the order, they tell us the exact value between units, AND they also have an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied.

Examples include height, weight, duration etc.

These data can be meaningfully added, subtracted, multiplied, divided (ratios).

Central tendency can be measured by mode, median, or mean; measures of dispersion, such as standard deviation and coefficient of variation can also be calculated from ratio scales.

Summary of Data Types

Provides:	Nominal	Ordinal	Interval	Ratio
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode, Median		✔	✔	✔
The "order" of values is known		✔	✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

Assignment 1

Redesign the infographic that accompanies *The Hindu* (Jan 14, 2016) article, **'Increase farmers' income, not production'.**