

You Need to Know Data: How to Become Data Literate in 7 Steps

RJ Williams
Senior Product Marketing Specialist
ASPE



Agenda

- Review core data terms and concepts
- Examine practical examples of data analysis and analytics for business
- Learn how you can help create a data-driven corporate culture

What Does Data Literacy Mean?

- Understanding what data means
- How to read graphs and charts correctly
- Draw accurate conclusions from data
- Recognize when data is being used in misleading or inappropriate ways.

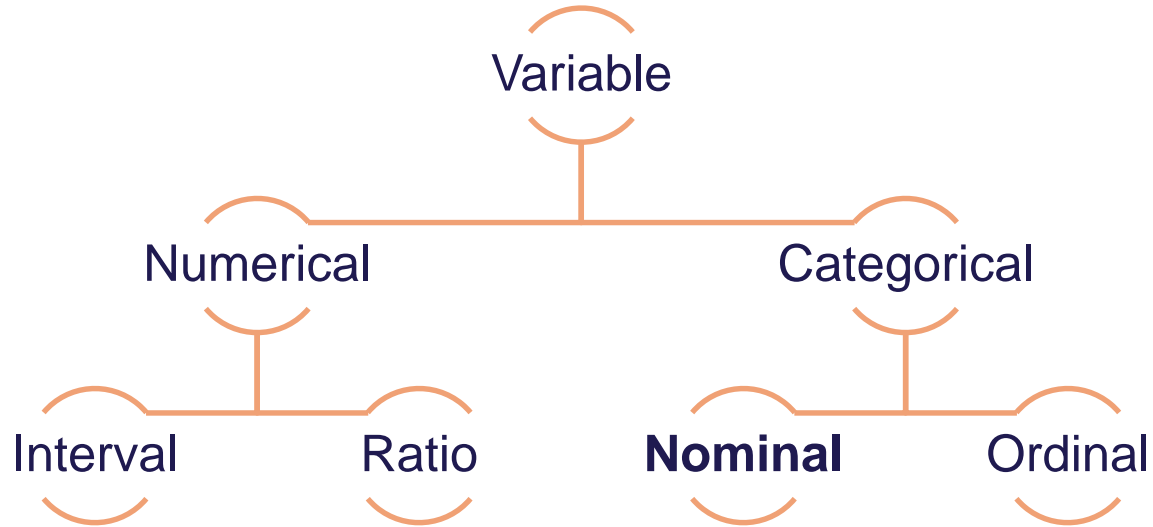
What are the 7 Steps?

You need to know:

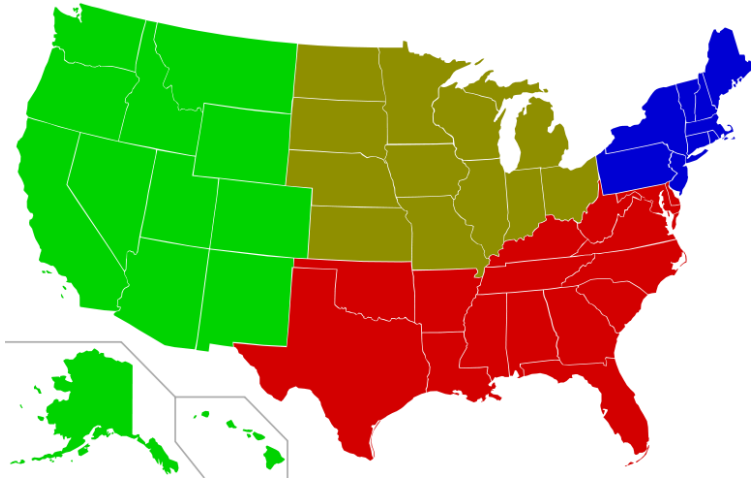
1. Types and attributes of data
2. Central tendency of data types and when to use them
3. How probability affects business
4. Types and characteristics of distributions
5. Variance
6. The difference between correlation and causation
7. How to combat cultural and cognitive barriers

Step 1: Types and Attributes of Data

Types of Variables



Nominal Variables



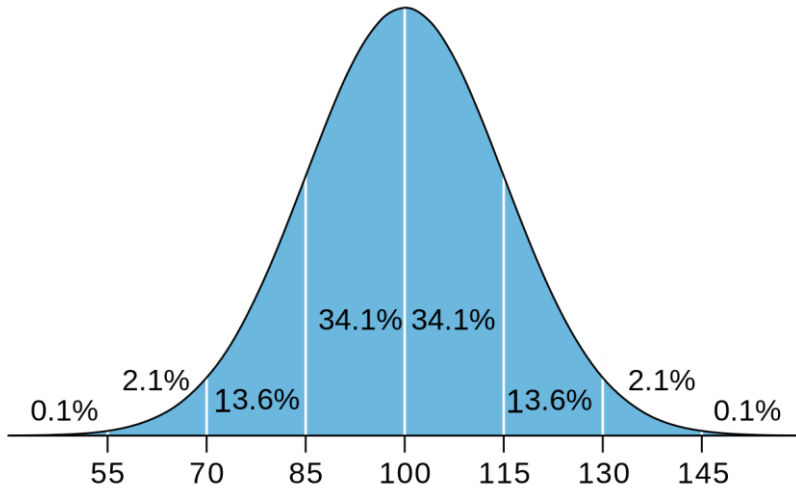
- Nominal = “Name ONLY”
- Nominal variables contain mere codes assigned to objects as labels, they are not measurements.
- Not a measure of quantity. Measures identity and difference.
- People either belong to a group or they do not.

Ordinal Variables: Order Matters

- Ranks individual attributes in same group
- Unit of measure not available
- Designates an ordering: greater than, less than
- Does not assume that the intervals between numbers are equal

2018 Rank	US. Business School
1	Stanford
2	Pennsylvania (Wharton)
3	Harvard
4	MIT (Sloan)
5	Chicago (Booth)
6	UC at Berkeley (Haas)
7	Columbia
8	Northwestern (Kellogg)
9	Virginia (Darden)
10	Cornell (Johnson)
11	Yale
12	Carnegie Mellon (Tepper)
13	NYU (Stern)
13	USC (Marshall)
15	Duke (Fuqua)
16	Washington (Foster)
17	UCLA (Anderson)
18	Michigan (Ross)
19	Dartmouth (Tuck)
20	Georgetown (McDonough)

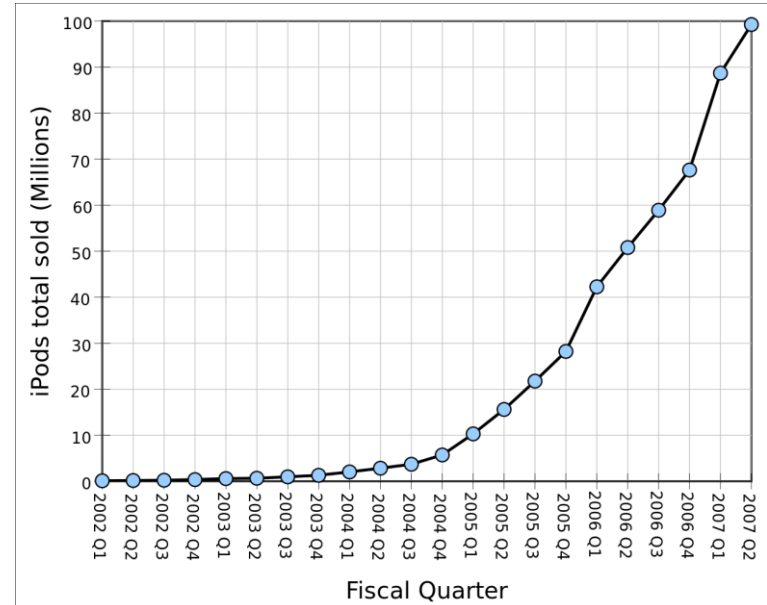
Interval Variables



- Classifies data into groups or categories
- Determines the preferences between items
- Zero point on the internal scale is arbitrary zero, it is not the true zero point
- Designates an equal-interval ordering.

Ratio Variables

- This is the highest level of measurement and has the properties of an interval scale; coupled with fixed origin or zero point.
- It clearly defines the magnitude or value of difference between two individual items or intervals in same group.



Summary of Variable Types

Level of Measurement (Variables)	Put Data in Categories	Arrange Data in Order	Subtract Data Values	Determine if one data value is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

Step 2: Central Tendency of Data

Central Tendency (Averages)

- Characterizes a set of data by a central or typical number
- Mean, Median, and Mode are most common measures of central tendency
- Each of them has strengths and weaknesses

(Arithmetic) Mean

- Most common meaning of “Average”
- Sum of all individual elements divided by the total number of elements
- Affected by sensitive to outliers
- Be careful about weights and units

(Arithmetic) Mean

Given a data set of: 1,2,3,3,4,11

$$\text{Mean} = \frac{1+2+3+3+4+11}{6} = 4$$

Median

- The midpoint of a distribution
- Half of the data above, half below
- Insensitive to outliers
- In the case of an even number of data points, take the mean of the middle two points

$$\text{Median}(1,2,3,3,4,11) = \frac{3 + 3}{2} = 3$$

Median

Median Individual Gross Weekly Earnings



Mode

- The value which appears most in a distribution
- A distribution may have more than one mode

$$\text{Mode}(1,2,3,3,4,11) = 3$$

Step 3: Probability

Probability

- The measure of the likelihood that an event will occur.
- Often associated with gambling, but has applications in business, finance, manufacturing, science and other fields.
- Generally expressed as a number between 0 and 1, where 0 means no chance of occurring and 1 means it will certainly occur.

Probability Terms

Event

The outcome to which the probability is assigned.

- “Heads” on a flipped coin.
- 4 on a rolled die
- A Queen drawn from a deck of cards

Sample Space

The set of all possible outcomes

- Heads and Tails
- 1, 2, 3, 4, 5, 6
- The 52 cards

Independent

The outcome of one event does not effect a different event

- Flipping two coins

Probability Uses In Business

- Data-driven companies apply the understanding of uncertainty and probability in their decision making process.
- Probability models can greatly help businesses in optimizing their practices and making better/safer decisions.
- Probability methods can increase the profitability and success of a business.

Probability In Business – Sales Predictions

- Anticipating future sales incomes is a popular application of probability.
- Companies use sales forecasts to predict revenues
- The probability distribution allows companies to see the worst and best possible outcomes and plan for both.

Probability In Business – Risk Assessments

- Probability distributions can help companies avoid negative outcomes
- Statistical analysis ventures that involve substantial risks
- The distribution shows which outcomes are most likely in a risky proposition and whether the rewards for taking specific actions compensate for those risks.

Step 4: Types and Characteristics of Distributions

Frequency Distribution

- The frequency with which observations are assigned to each category or point on a measurement scale.
- Most basic form of descriptive statistics
- May be expressed as a percentage of the total sample found in each category

Frequency Distribution

How the distribution is read depends on the measurement level.

- Nominal scales are read as discrete measurements at each level
- Ordinal measures show tendencies, but categories should not be compared
- Interval and ratio scales allow for comparison among categories

Frequency Distribution

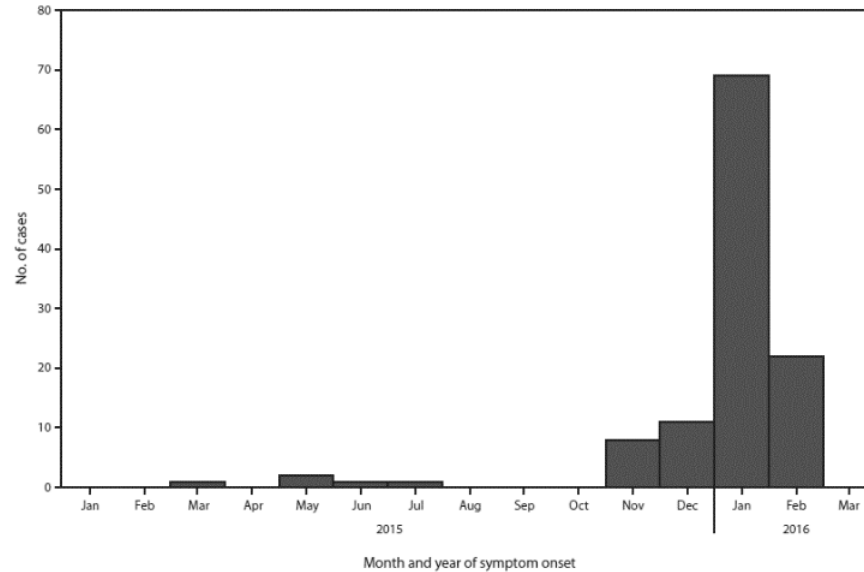


Figure. Month of illness onset for 115 patients with laboratory evidence of Zika virus infection among residents of U.S. states and the District of Columbia — January 1, 2015–February 26, 2016*

Step 5: Types of Variances

Measure and Variability

- There will always be **variability** in the data.
- Statistics helps measure and characterize variability.
- For example, controlling (or reducing) variability in a manufacturing process equates to statistical process control.

Range

- The difference between the greatest and least data point
- Useful, but susceptible to outliers

$$\text{Range}(1,2,3,3,4,11) = 11 - 1 = 10$$

Quartiles

- Divide distribution into 4 sets with equal numbers of points
- 25% of the data is less than the First (or Lower) Quartile
- The Median is the Second Quartile
- 25% of the data is greater than the Third (or Upper) Quartile
- The Interquartile Range is the difference between the Third and First Quartiles

Variance

- Squared differences between the mean and each observation, divided by N
- It is positive but not scaled to the observations
- Say, miles squared versus miles

Standard Deviation

- Used to quantify and illustrate the amount of disparity in a set of data values
- The square root of the variance
- It is scaled to the observations and is also positive
- Low value indicates values are close together
- High value indicates values are spread over wider range

Step 6: Correlation and Causation

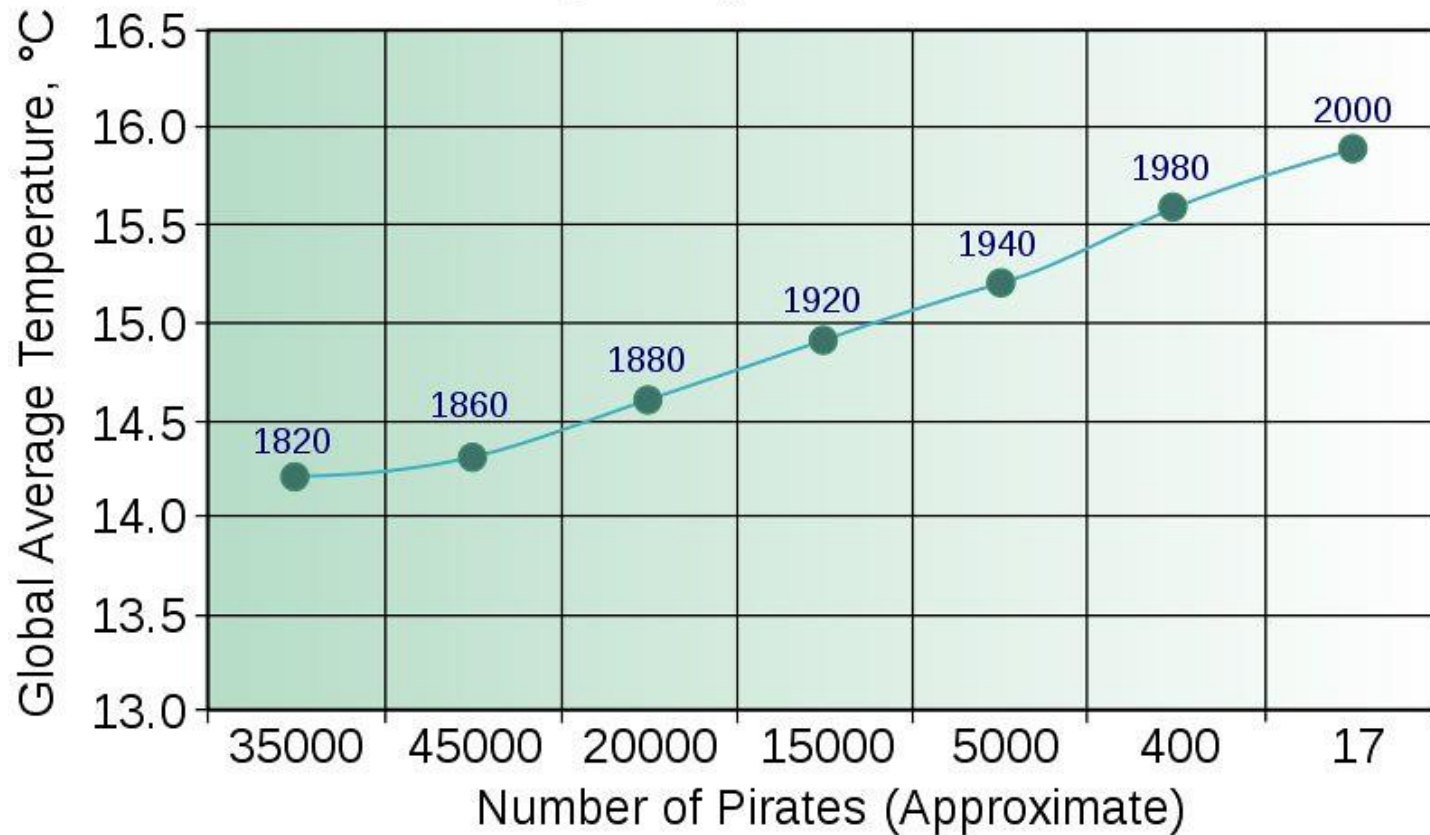
Correlation and Causation

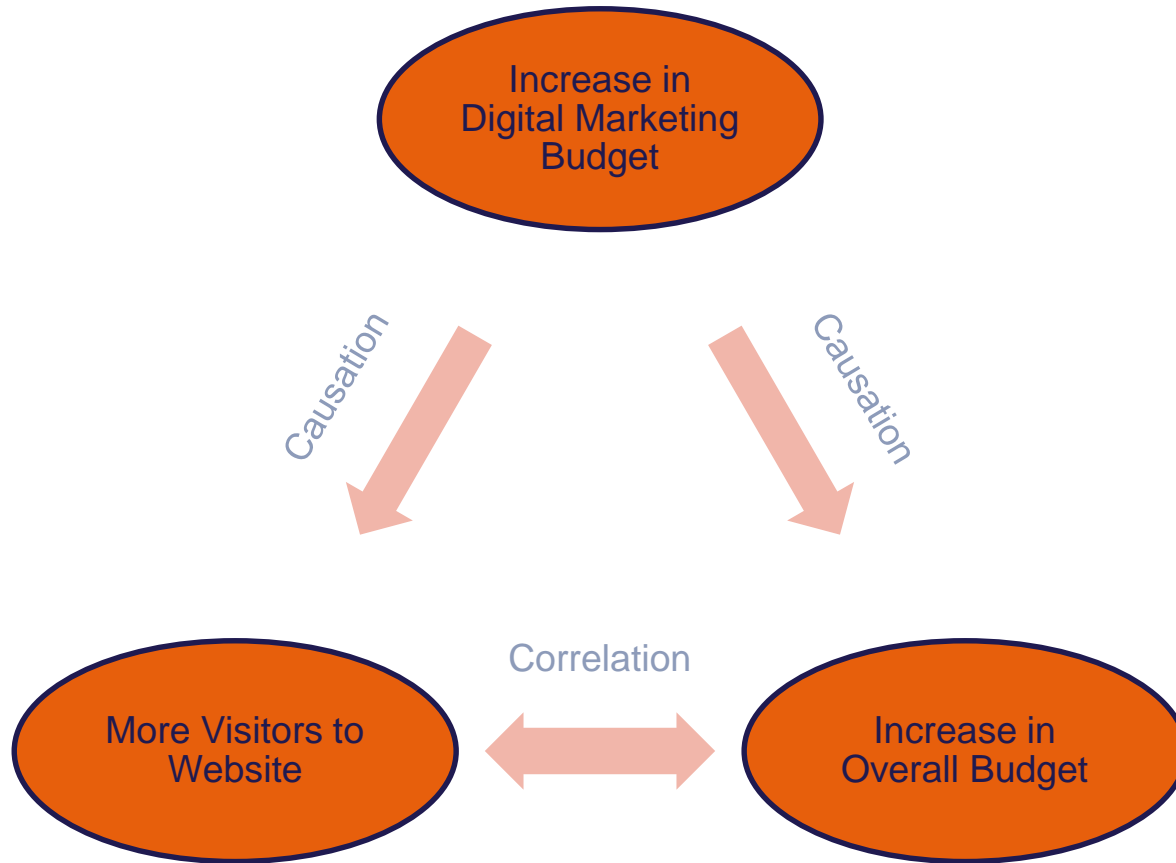
Two variables that change in similar ways, perhaps because they are linked to something else.

VS

An independent factor that directly impacts a dependent one.

Global Average Temperature vs. Number of Pirates





Correlation and Causation

- Knowing the difference between the two goes a long way in ensuring that your business decisions are based on hard facts and measurable variables.

Step 7: Creating a Data-Driven Culture

**ONE ACCURATE
MEASUREMENT IS WORTH A
THOUSAND EXPERT
OPINIONS.**

– GRACE HOPPER

Barriers to Data-Driven Decisions

Cultural

- HiPPO
- The Value of Intuition
- Lack Accountability

Cognitive

- Confirmation Bias
 - Recency Bias
 - Illusion of Validity
- Rigid Mental Models

Solutions to Barriers

Incentives and
Accountability

Data Based Proof

Transparency

Ability

Actions Tied to Outcomes

- Performance is tied to quantitative measures.
- Metrics will reflect poor decisions and good ones as well.

Solutions to Barriers

Incentives and
Accountability

Data Based Proof

Transparency

Ability

Actions Tied to Outcomes

- Move from instinct to proof based culture in the organization.
- Analyze results of testing (like A/B for user interface options).
- Use the evidence.

Solutions to Barriers

Incentives and
Accountability

Data Based Proof

Transparency

Ability

Actions Tied to Outcomes

- Data driven culture is transparent.
- Responsibilities are clear.
- Decisions are made, outcomes are known.
- Evidence is available and accountability is high.

Solutions to Barriers

Incentives and
Accountability

Data Based Proof

Transparency

Ability

Actions Tied to Outcomes

- Positively affect people's ability to perform their tasks.
- Providing the money, the right time frame, the right expectations lowers the barriers to making good decisions.

Solutions to Barriers

Incentives and
Accountability

Data Based Proof

Transparency

Ability

Actions Tied to Outcomes

- The business need dictates the problems to be solved.
- The problems dictate the data analyses.
- The results should be presented in a way that ties everything back to the business need (recommendations and impact).

Questions?

Thank You!

- **Data Literacy for Business Professionals eLearning**
<https://aspetraining.com/courses/data-literacy-training-for-business>
- **Introduction to Data Analysis**
<https://aspetraining.com/courses/introduction-to-data-analysis>
- **Data Analysis Boot Camp**
<https://aspetraining.com/courses/data-analysis-boot-camp>