# Business Analytics and Intelligence

1

## Agenda

- Business Intelligence Foundation
- Descriptive and Diagnostic Analytics
- Predictive Analytics
- Storytelling with Data

2

Business

Intelligence

Foundation

3

# Business Intelligence

4

# Data Driven Decision Making

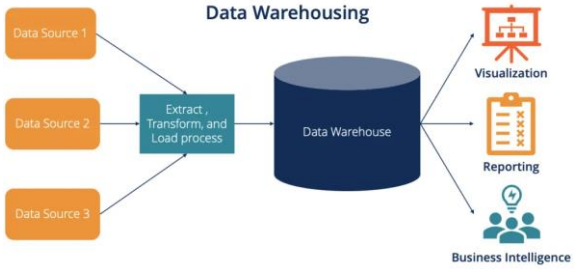
- The practice of basing decisions on the analysis of data, rather than purely on intuition.
- Research showed that Data Driven Decision Making (DDD) was associated with higher productivity and market value.
- usage of data for the creation of a new product or service.
- usage of data for business decision making in the entire company.





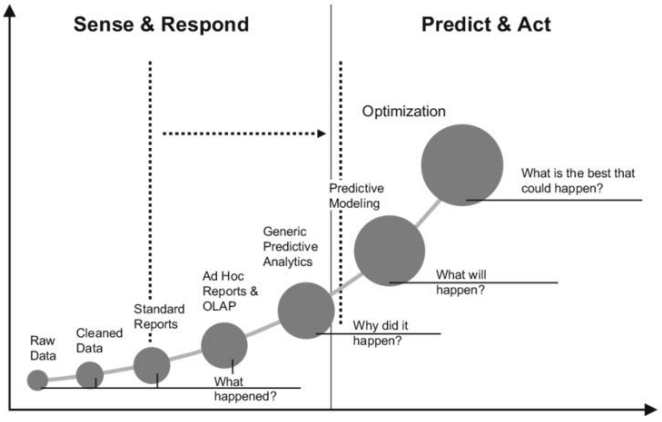
# Evolution of Database Technology

- Files
- Database
- Data Warehouse
- Data Lake




Credit : https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/

# Evolution of Analytics




Credit : https://www.researchgate.net/figure/Evolution-of-mature-analytics-capabilities-adapted-from-SAP-2012_fig5_311998130





# Business Intelligence and Data Mining

- Business intelligence (BI) : An umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance.

  *** Analysis and Reporting ***

- Data Mining : The process of discovering meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories. Data mining employs pattern recognition technologies, as well as statistical and mathematical techniques.

  *** Analysis and Modeling ***

# Business Analytics

- Business analytics is the process of transforming data into insights to improve business decisions. Data management, data visualization, predictive modeling, data mining, forecasting simulation, and optimization are some of the tools used to create insights from data.

  - identifying new patterns and relationships with data mining.

  - using quantitative and statistical analysis to design business models.

  - conducting A/B and multi-variable testing based on findings;

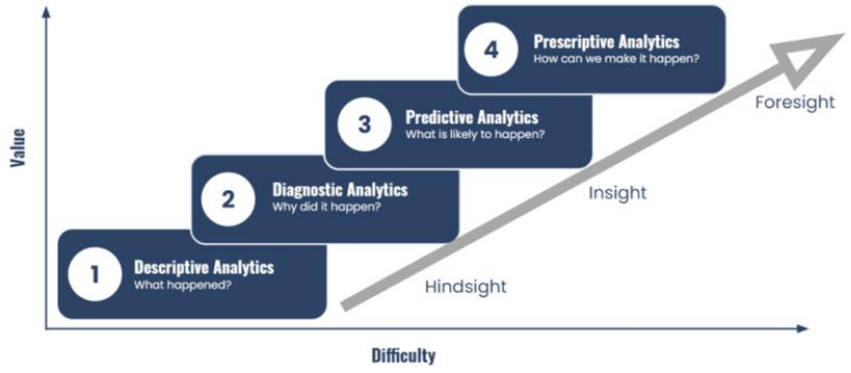
  - forecasting future business needs, performance, and industry trends with predictive modeling.

  - communicating your findings in easy-to-digest reports to colleagues, management and customers.





# Business Analytics Maturity




Credit : https://www.phdata.io/blog/what-is-analytics-maturity-framework/

## Data Analytics Lifecycle


Credit : https://medium.com/@shawn.chumbar/the-crisp-dm-process-a-comprehensive-guide-4d893aecb151



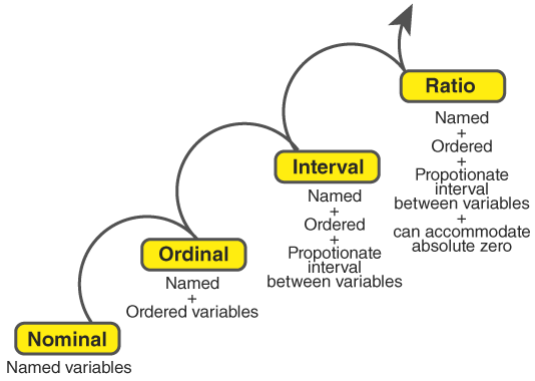


# Descriptive and Diagnostic Analytics

# Measurement

13

## Level of Measurement




Credit : https://byjus.com/maths/scales-of-measurement/

14

14

# Nominal

- A nominal scale is the 1st level of measurement scale in which the numbers serve as “tags” or “labels” to classify or identify the objects.
- A nominal scale usually deals with the non-numeric variables or the numbers that do not have any value.
- With nominal data, as the name implies, the numbers function as a name or label and do not have numeric meaning.
- Variables with nominal scale

  - Gender

  - Marital Status

Credit : https://byjus.com/ & Dr. Thanachart Ritbumroong

15



# Ordinal

- The ordinal scale is the 2nd level of measurement that reports the ordering and ranking of data without establishing the degree of variation between them. Ordinal represents the “order.”
- Ordinal data is known as qualitative data or categorical data. It can be grouped, named and also ranked.
- Ordinal data refers to data that has some meaningful order, so that higher values represent more of some characteristic than lower values.
- Variables with ordinal scale

  - Exam grade

  - Team ranking

Credit : https://byjus.com/ & Dr. Thanachart Ritbumroong

16

# Interval

- The interval scale is the 3rd level of measurement scale. It is defined as a quantitative measurement scale in which the difference between the two variables is meaningful.
- In other words, the variables are measured in an exact manner, not as in a relative way in which the presence of zero is arbitrary.
- Interval data has a meaningful order and has the quality of equal intervals between measurements, representing equal changes in the quantity of whatever is being measured.
- Variables with interval scale

  - Fahrenheit temperature (Fahrenheit scale has no natural zero point because 0 on the Fahrenheit scale does not represent an absence of temperature but simply a location relative to other temperatures)
  - IQ test score (lowest = 40)

Credit : https://byjus.com/ & Dr. Thanachart Ritbumroong

17



# Ratio

- The ratio scale is the 4th level of measurement scale, which is quantitative. It is a type of variable measurement scale. It allows researchers to compare the differences or intervals. The ratio scale has a unique feature. It possesses the character of the origin or zero points.
- Ratio data has all the qualities of interval data (meaningful order, equal intervals) and a natural zero point.
- Variables with ratio scale

  - Age
  - Weight
  - Income

Credit : https://byjus.com/ & Dr. Thanachart Ritbumroong

18

# Categorical and Numerical

19

## Proxy Measurement

- The term proxy measurement refers to the process of substituting one measurement for another.
- How would you measure customer loyalty?

Credit : Dr. Thanachart Ritbumroong

## True and Error Scores

- Classical measurement theory conceives of any measurement or observed score as consisting of two parts: true score (T) and error (E). This is expressed in the following formula:

$$X = T + E$$

- Where X is the observed measurement, T is the true score, and E is the error.

Credit : Dr. Thanachart Ritbumroong

21



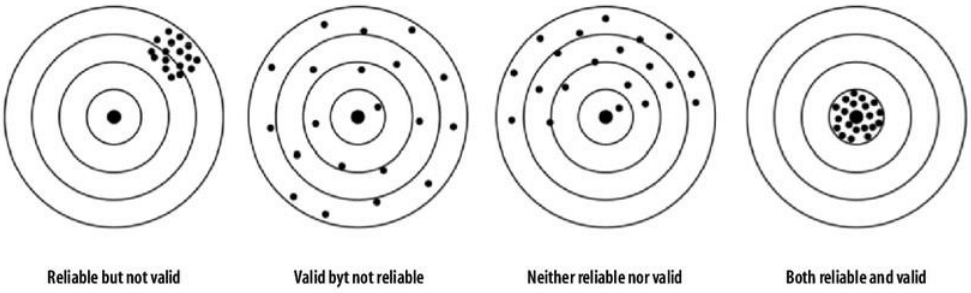
## Random and Systematic Error

- Because we live in the real world rather than a Platonic universe, we assume that all measurements contain some error.
- However, not all error is created equal, and we can learn to live with random error while doing whatever we can to avoid systematic error.
- Random error is error due to chance.
- Systematic error has an observable pattern.

Credit : Dr. Thanachart Ritbumroong

22

## Reliability and Validity




Credit : Dr. Thanachart Ritbumroong





## Measurement Bias

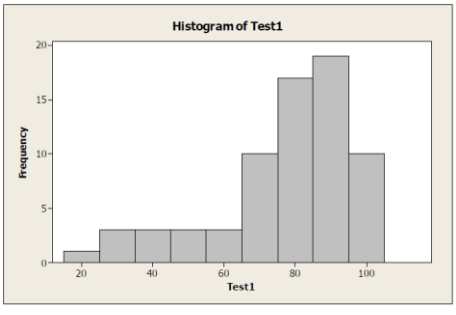
- Our product is priced below that of other companies on the markSelection bias
- Volunteer bias
- Recall bias
- Social desirability biaset

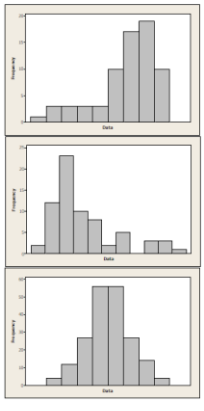
Credit : Dr. Thanachart Ritbumroong

# Descriptive Statistics

25

## Describing Quantitative Data - Histograms



Skewed to the left

Skewed to the right

Symmetric

Credit : Dr. Thanachart Ritbumroong

26

26

## Measures of Localization

- Quartile
- Percentile
- Deciles

Credit : Dr. Thanachart Ritbumroong





## Measures of Centrality

- Mode
- Median
- Mean

Credit : Dr. Thanachart Ritbumroong

# Measures of Variation

- Range
- Variance
- Standard deviation
- Coefficient of variance

| | Range | Standard deviation |
|---|---|---|
| Nominal | No | No |
| Ordinal | Yes (NOT the best method) | No |
| Interval Ratio | Yes (NOT the best method) | Yes (if data is symmetric and unimodal) |

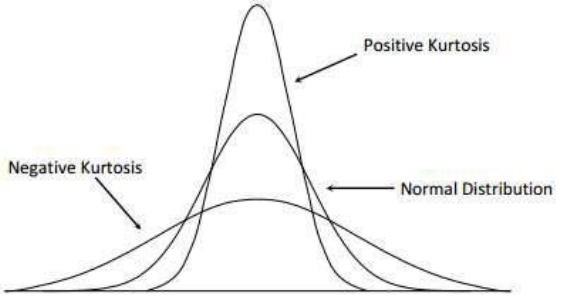
Credit : Dr. Thanachart Ritbumroong

29

29

# Measures of Symmetry

- Skewness
- Kurtosis




Credit : Dr. Thanachart Ritbumroong

30

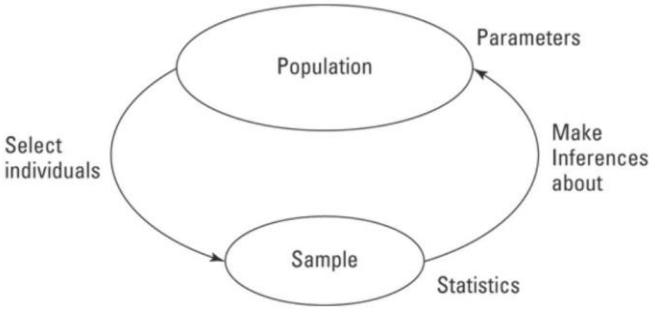
30

# Hypothesis Testing



## Samples and populations

- In a nutshell, statistics is all about — using the data from samples to draw conclusions about populations.




Credit : Dr. Thanachart Ritbumroong

## What is a Hypothesis?

- A hypothesis is a claim (assumption) about a population parameter:
- population mean : The mean monthly cell phone bill of this city is = $42
- population proportion : The proportion of adults in this city with cell phones is p = .68

Credit : Dr. Thanachart Ritbumroong

33



## Hypothesis Testing Process

- State the hypotheses
- Collect data
- Calculate the test statistic
- Draw conclusions

Credit : Dr. Thanachart Ritbumroong

34

# Errors in Making Decisions

- Type I Error :

  Reject a true null hypothesis

  Considered a serious type of error

- The probability of Type I Error is :

  Called level of significance of the test

  Set by researcher in advance

- Type II Error

  Fail to reject a false null hypothesis

Credit : Dr. Thanachart Ritbumroong





# Outcomes and Probabilities

**Possible Hypothesis Test Outcomes**

| | State of Nature | |
|---|---|---|
| **Decision** | $H_0$ True | $H_0$ False |
| Accept $H_0$ | **No error** $(1 - \alpha)$ | **Type II Error** $(\beta)$ *False Negative* |
| Reject $H_0$ | **Type I Error** $(\alpha)$ *False Positive* | **No Error** $(1 - \beta)$ |

**Key:**
**Outcome**
**(Probability)**

Credit : Dr. Thanachart Ritbumroong

# Mean Comparison

37

## Mean Comparison

- Two random samples are drawn from the two populations of interest.
- Examples of independent samples;
  - Mean difference between male and female
  - Mean difference between section 1 and section 2
  - Mean difference between Thailand and Singapore


Credit : Dr. Thanachart Ritbumroong

38

38

# ANOVA

39

## ANOVA

- Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.
- Two variables: 1 Categorical, 1 Quantitative
- Main Question: Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?
- If categorical variable has only 2 values:
  - 2-sample t-test
- ANOVA allows for 3 or more groups

Credit : Dr. Thanachart Ritbumroong

40

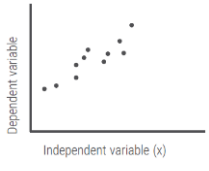
40

# Regression

41

## Regression

- Regression is the attempt to explain the variation in a dependent variable using the variation in independent variables.
- Regression is thus an explanation of causation.
- If the independent variable(s) sufficiently explain the variation in the dependent variable, the model can be used for prediction.




Credit : Dr. Thanachart Ritbumroong

42

42

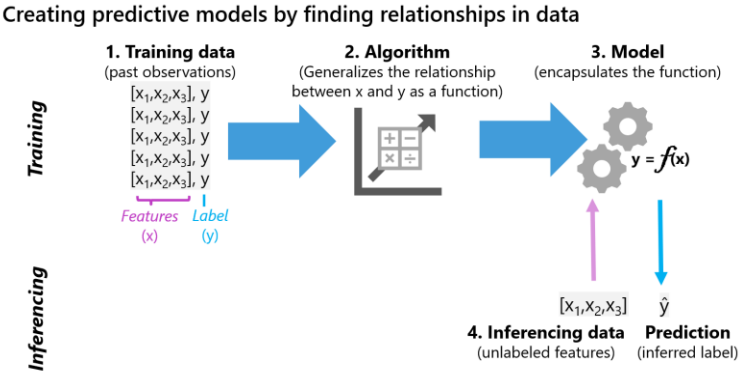
# Predictive Analytics





# Understanding Machine Learning

# Understanding Machine Learning

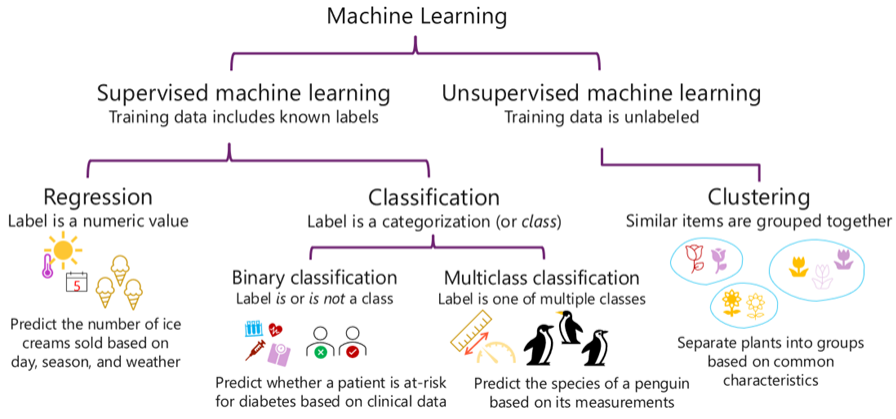



Credit : Microsoft Learn





# Understanding Machine Learning




Credit : Microsoft Learn

# Supervised Learning

47

## Supervised Learning

- **Regression** algorithms are used to predict a continuous numerical output. For example, a regression algorithm could be used to predict the price of a house based on its size, location, and other features.

- **Classification** algorithms are used to predict a categorical output. For example, a classification algorithm could be used to predict whether an email is spam or not.

Credit : https://www.geeksforgeeks.org/



48

# Unsupervised Learning

49

## Unsupervised Learning (Clustering )

- The task of grouping data points based on their similarity with each other is called Clustering or Cluster Analysis.
- This method is defined under the branch of Unsupervised Learning, which aims at gaining insights from unlabelled data points, that is, unlike supervised learning we don’t have a target variable.

Credit : https://www.geeksforgeeks.org/



50

# Reinforcement Learning

51

## Reinforcement Learning

- Reinforcement Learning (RL) is a branch of machine learning focused on making decisions to maximize cumulative rewards in a given situation.
- Unlike supervised learning, which relies on a training dataset with predefined answers, RL involves learning through experience.
- In RL, an agent learns to achieve a goal in an uncertain, potentially complex environment by performing actions and receiving feedback through rewards or penalties.

Credit : https://www.geeksforgeeks.org/

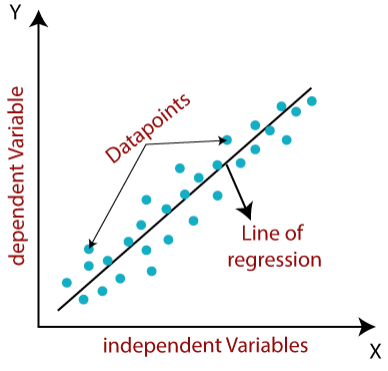


52

# Algorithms in Machine Learning

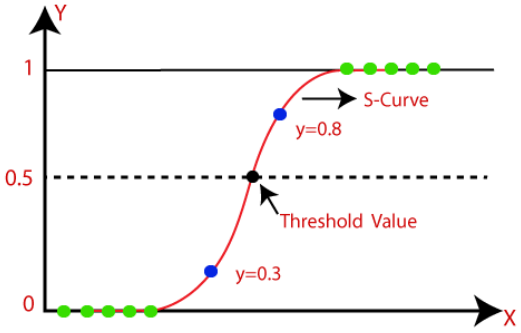


## Linear Regression




Credit : https://www.javatpoint.com/linear-regression-in-machine-learning

# Logistic Regression

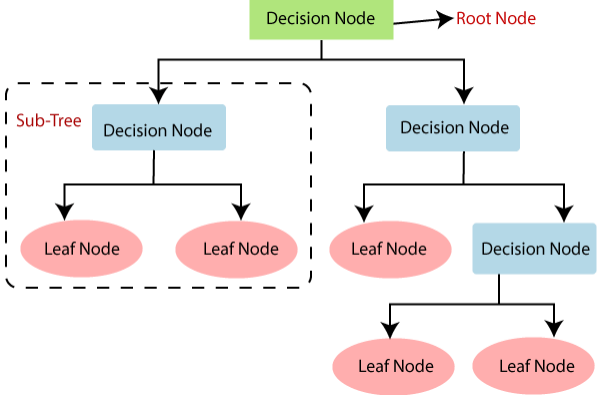



Credit : https://www.javatpoint.com/linear-regression-in-machine-learning

55

55

# Decision Trees




Credit : https://www.javatpoint.com/linear-regression-in-machine-learning

56

56

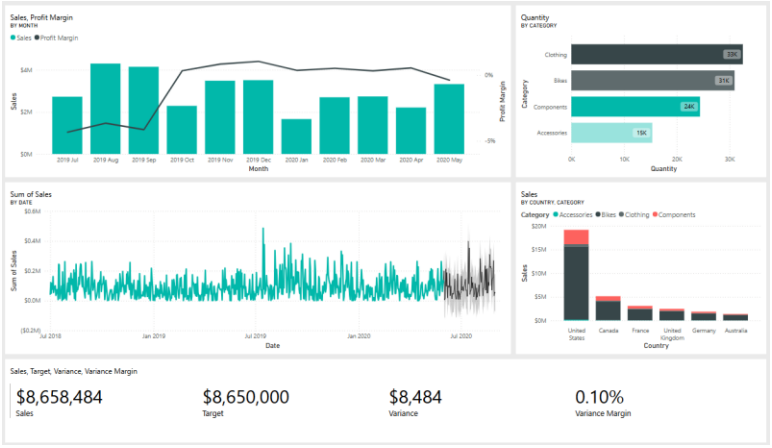

# Storytelling with Data



## Dashboard Design Principle

- Use a familiar chart type
- Add no more than 5 slices to a pie chart (Recommend 3 slices)
- Order the data series
- Avoid 3D charts
- Don’t use randomly generated colors
- Don’t include a legend when it’s not needed
- Only use grid lines when it’s helpful

# Visual Analytic








# Storytelling

Context

Visuals

Clutter

Attention

Design

Story

Credit :storytelling with data - Cole Nussbaumer Knaflic @Wiley

## Context

Who :
To whom are you communicating?

What :
What do you want your audience to know or do?

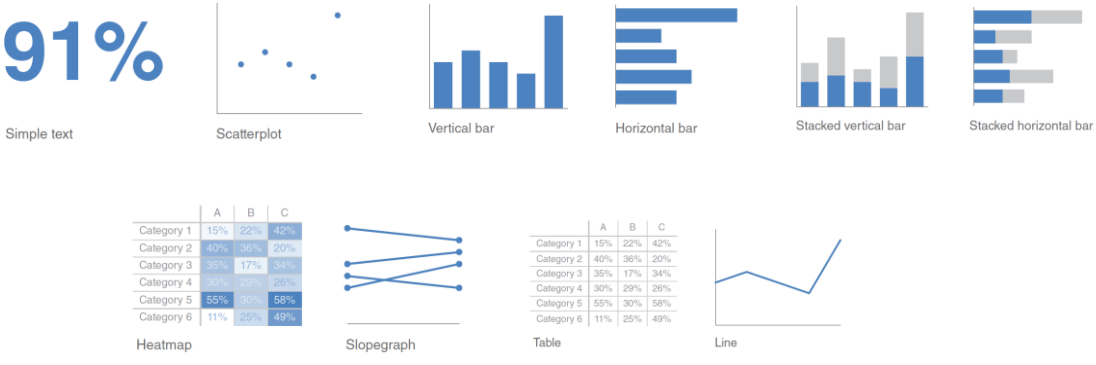
How :
How can I make data to support my point?

Credit :storytelling with data - Cole Nussbaumer Knaflic @Wiley

61

61

## Effective Visual




Credit :storytelling with data - Cole Nussbaumer Knaflic @Wiley

62

62

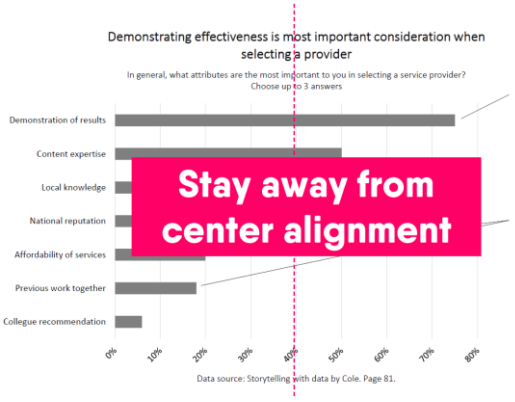
# Clutter

Contrast

Repetition

Alignment

Proximity




Credit :storytelling with data - Cole Nussbaumer Knaflic @Wiley & Apirak Panatkool - Founder of UX Academy





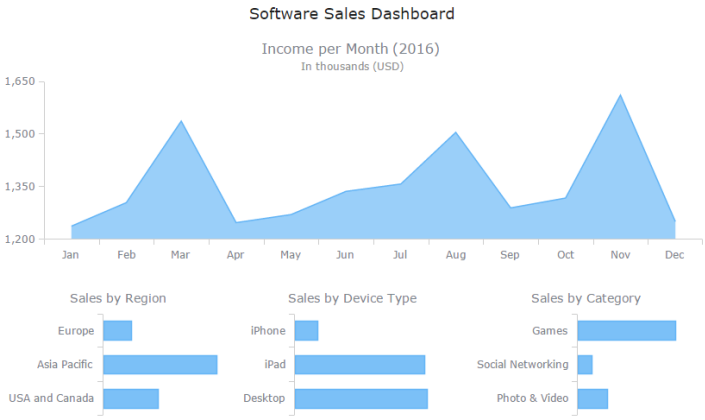
# Attention

We can enable our audience to see what we want them to see before they even know they're seeing it.

Credit :storytelling with data - Cole Nussbaumer Knaflic @Wiley

# Design

- Affordance
- Accessibility
- Aesthetics
- Acceptance

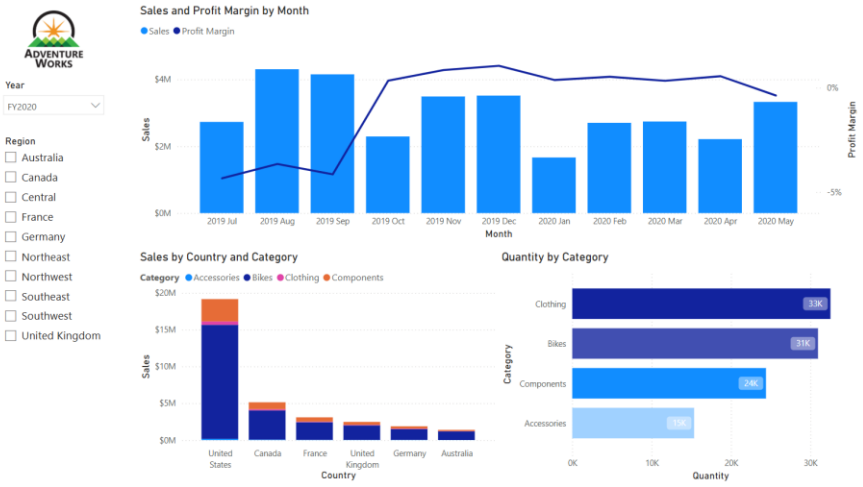



Credit : https://www.anychart.com/blog/2017/04/05/chart-captions-title-graph-tips/





# Story



Thank You