
Convey Your Message At a Glance

Aryeh Primus, Data Engineer

Agenda

- Measurement Types
- The Human Visual System
- Examples: 2 Variables
- Example: 3 Variables
- Q&A

What is a Measurement

- At the simplest level, a measurement is an answer.
- Common usage:
 - If you use a tape measure to find how tall your desk is, the height is a measurement.
 - If you use a stop watch to time how long it takes to run a lap, that's a measurement.
 - The number of open request tickets is a measurement.
 - The response rate to an advertising campaign is a measurement.
- But there's more – measurements don't have to be objective
 - If you send out a CSAT survey, the answer to each question is a measurement.
 - If you ask someone their favorite color, that's a measurement.

Measurement Types

- Not all measurements are equal
- How we can use the measurement depends on its type
- All measurements fall into one of four Types (or Scales):
 - Nominal
 - Ordinal
 - Metric
 - Ratio
- *The key to knowing which visualizations are the best is understanding what types of measurements you're showing.*

Nominal

- Nominal variables represent categories of things
- As an example, if we want to know someone's favorite fruit we could get values of Apple, Banana and Orange
- If we asked 1000 people their favorite fruit, and put the answers into Excel, there isn't much we can do with it
- In fact, the only thing we can do is show how many people liked each fruit

Nominal Operations: =, !=

Ordinal

- Ordinal variables represent things that can be put in order
- An example is "City Size", which could have values of "Town", "City", "Metropolis"
- We can say "Town" < "City" < "Metropolis"
 - Remember, we can't say "Apple" < "Orange"
- While we know that "Town" < "City", we don't know exactly how much bigger City is.
We just know that it's bigger

Ordinal Operations: =, !=, <, >

Metric (or Interval)

- Metric variables have defined distances between them.
- The classic example of a metric measure is a date.
 - We know that there are exactly two days between 05-May and 03-May
- We can say, for instance, "The time between 05-May and 03-May is less than the time between 20-May and 5-May"
- *Note: Not everything that's a number is a metric variable. Think about apartment numbers, zip codes, or jersey numbers on a sports team.*

Metric Operations: =, !=, <, >, +, -

Ratio

- Ratio variables have a meaningful zero value
- We can multiply and divide ratio variables
- Dates aren't a ratio measure because there's no real answer to "What is twice as big as 2-Apr?"
- Examples of ratio measures include
 - Physical distance -- We know what half six feet is. It's 3 feet.
 - Age -- We all know that a 30 year old is twice the age of a 15 year old.
 - Correct answers on a test, Days of work missed, Percent response rate...

Ratio Operations: =, !=, <, >, +, -, *, /

Summary of Types

Measure Type	Criteria	Operations	Example
Nominal	Categories	=, !=	Birth Country
Ordinal	Can be ordered	=, !=, <, >	CSAT Score
Metric / Interval	Have distance	=, !=, <, >, +, -	Dates
Ratio	Have a "0" value	=, !=, <, >, +, -, *, /	Height

Quiz – Question 1

The TSA wants to track how long people wait in the security lines. Assume they have a system where, when a passenger gets in the line, their boarding pass is scanned. Then, when a passenger finishes security, the boarding pass is scanned again. This gives the elapsed time from when a customer starts security until it's completed. What measurement type do they get?

- A) Nominal
- B) Ordinal
- C) Interval
- D) Ratio

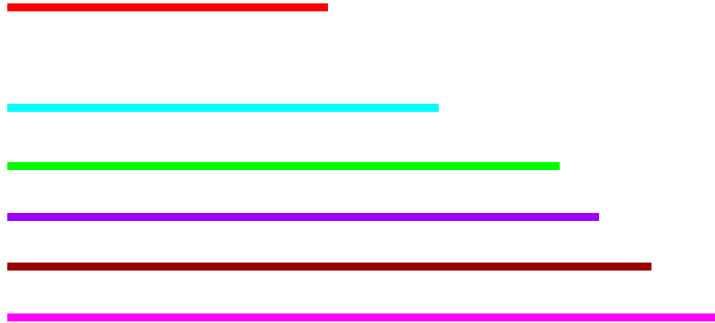
Quiz – Question 2

After completing security, there's a little box with three buttons. One button has a smiley face, one a neutral face, and one a frowny face. Passengers can push a button to indicate if they were happy, neutral, or unhappy with the security experience. What type of measure is this?

-
- A) Nominal
 - B) Ordinal
 - C) Interval
 - D) Ratio

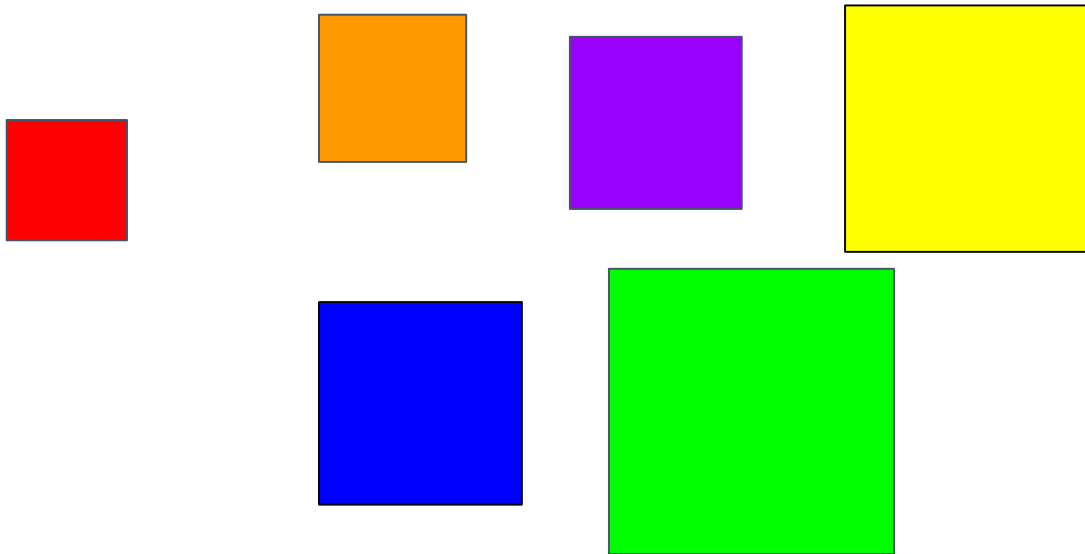
Visual System

Look at the line in red. Which other line is twice as long?



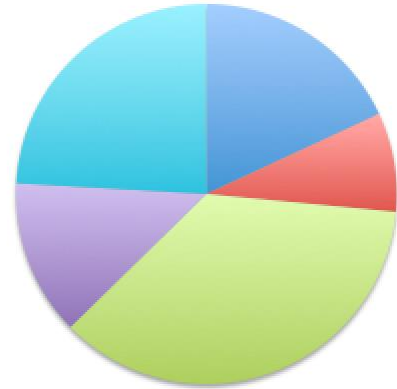
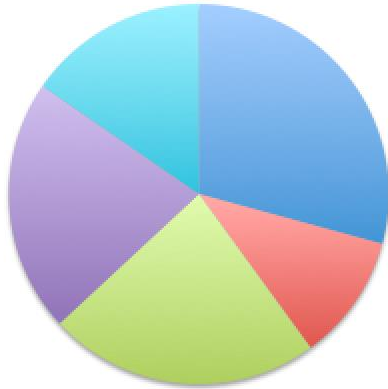
Visual System

Look at the box in red. Which other box is twice as big?



Visual System

Look at the following pie charts. In which one is the purple wedge twice as big as the red wedge?

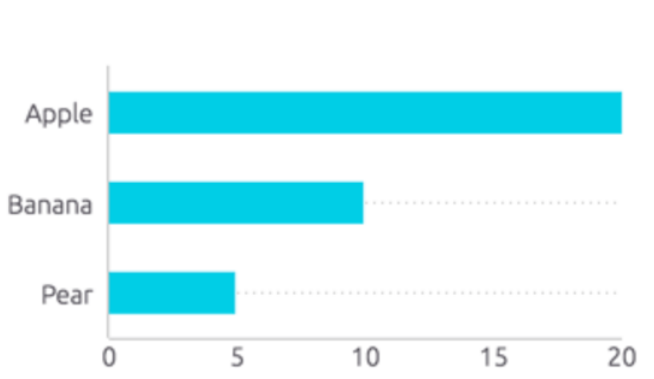
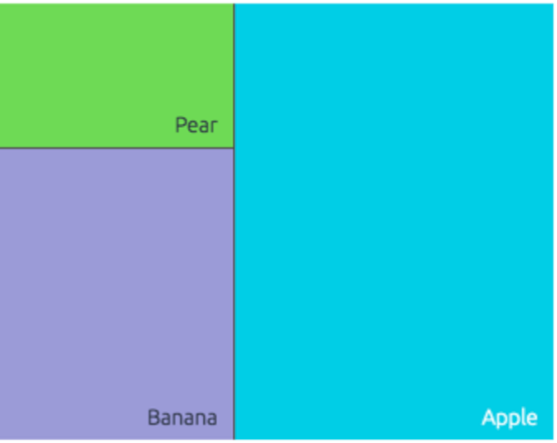
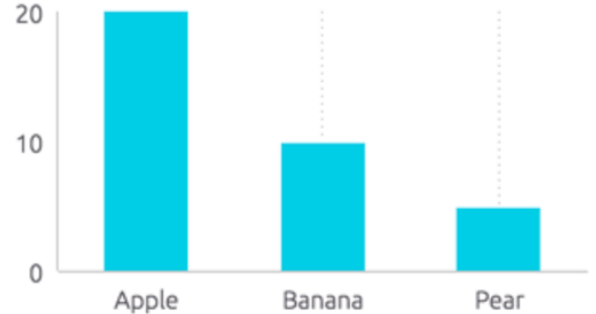
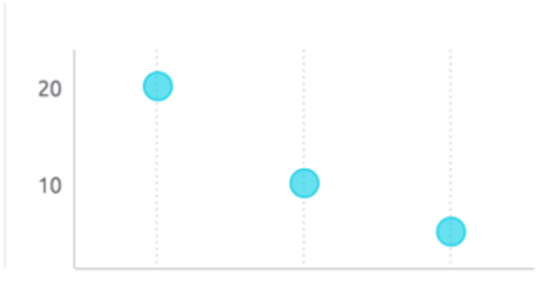
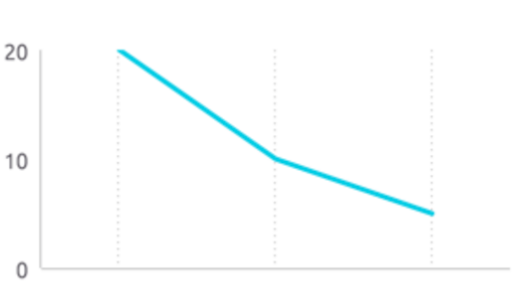


2 Variable Example A

A survey was taken where 35 people were asked for their favorite fruit. The breakdown of responses was as follows.

Fruit	# People
Apple	20
Banana	10
Pear	5

2 Variable Example A - Which is best?

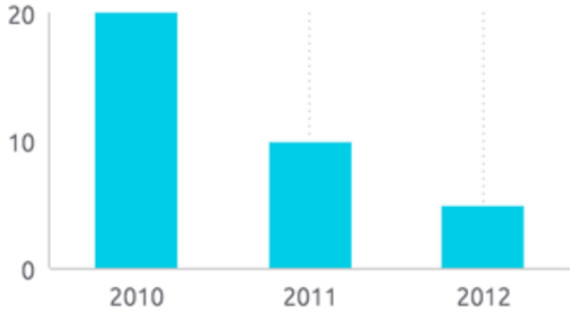
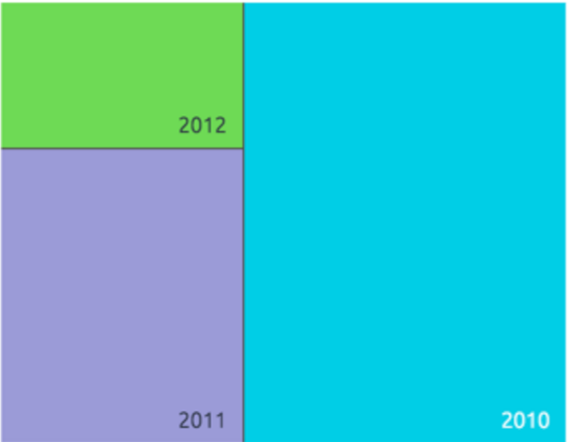
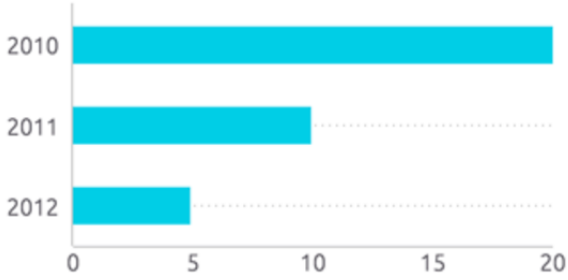


2 Variable Example B

Instead of a survey on favorite fruit, think about a company with the following yearly sales figures.

Year	Sales (\$M)
2012	5
2011	10
2010	20

2 Variable Example B - Which is best?



2 Variable Summary

Chart Type	Fruits (Nominal)	Years (Interval)
Line	Bad	Best
Scatter	Bad	OK
Treemap	Good	Bad
Pie	Good	Bad
Bar	Best	OK
Column	OK	Best

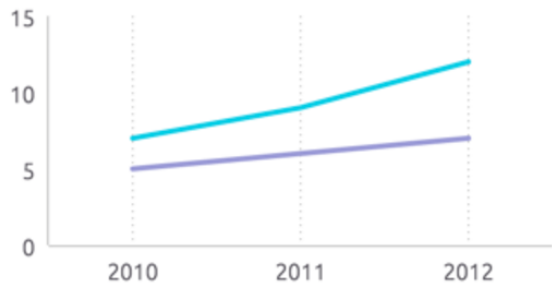
3 Variable Example

This time, let's consider data with more than one dimension. Think about a company that sells two types of products: shoes and household goods. Below is a table showing the sales over three years.

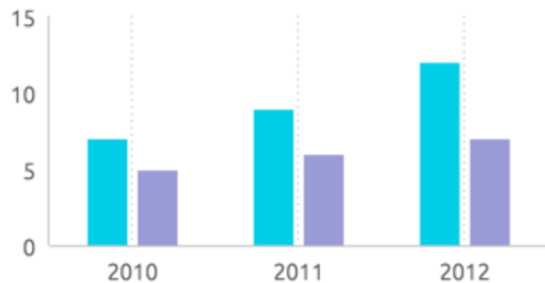
Year	Department	Sales
2010	Shoes	5
2010	Household	7
2011	Shoes	6
2011	Houshold	9
2012	Shoes	7
2012	Household	12

3 Variable Example - Options

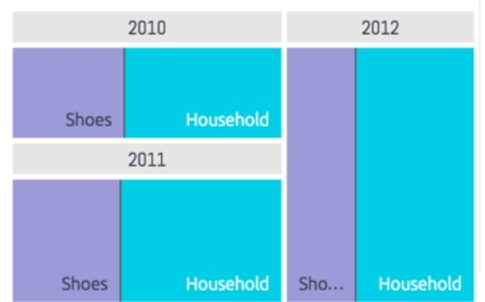
Line



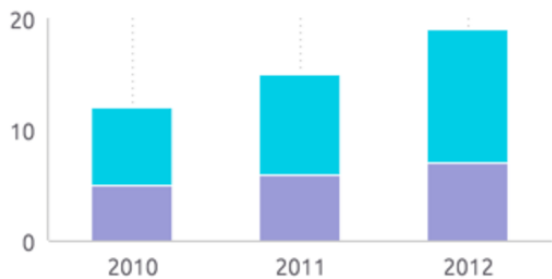
Side-by-Side Column



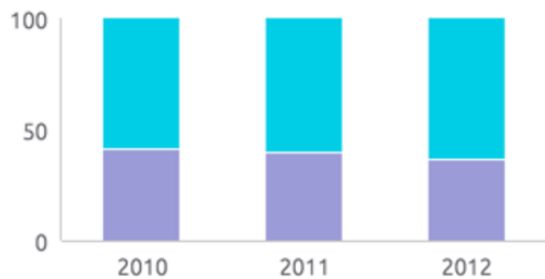
Treemap



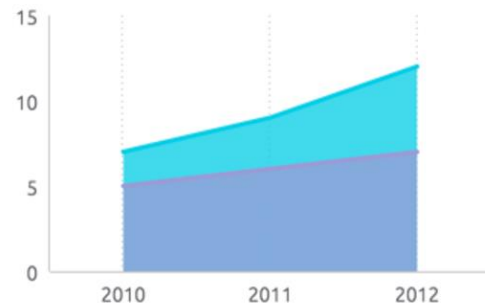
Stacked Column



100% Column



Area



3 Variable Example

So which is best? With the exception of the treemap, each of the charts could be the best option. Unlike in the first two examples, the choice here is largely about what you're trying to convey. What question is the widget answering?

- If the goal is to convey the fact that household is growing faster than shoes, and that the difference in speeds is increasing, the line chart is the winner.
- If the chart is supposed to show that the revenue mix is tilting more towards household goods over time, the 100% Column chart is best.
- The Stacked Column Chart is the only widget that clearly expresses total revenue over time, but it does so at the expense of the revenue breakdown.

Key Takeaways

- Always start by asking, “What message do I want to convey”?
- Think about the data you’re presenting and the type of each data.
- Don’t imply relationships that aren’t real.
- Don’t hide relationships that are real.
- Remember that Distance > Area > Arc