03. Data

ID 413: Information Graphics and Data Visualization
Spring 2025

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

Assignment 3: Comparative Visualization

One of the most important questions that a visualization attempts to answer is, "Compared to what?" Comparisons are at the heart of understanding quantitative information.

To help with easy comparison, we do one or more of the following:

1. Translate
2. Transpose
3. Transform

Assignment 3: Comparative Visualization

Translate -- Apply appropriate measurement scales to comprehend the data. Raw numbers to percentages, time into years, months, days, hours, minutes or seconds, lengths/distances into cms, mts or kms, areas into sq.units, volumes, weights, temperatures, and so on.

Assignment 3: Comparative Visualization

Assignment 3: Comparative Visualization

Transform -- Convert measurement scales into something more suitable for the visualization goal. Make it familiar & relatable to the reader.

Filtering/Refining: Removing irrelevant data points. Replacing codes with descriptive labels.

Aggregation: Combining data points into summaries like averages or totals

Normalization: Scaling data to a common range. Negative & log scales.

Binning: Grouping data into discrete ranges. For example, grouping age ranges into "young adult," "middle-aged," and "senior".

Data type conversion: Changing data types (e.g., text to numeric)

Assignment 3: Comparative Visualization

Transform -- Convert measurement scales into something more familiar & relatable to the reader.

TERMS OF SERVICE

HOW LONG IT WOULD TAKE TO READ THE TERMS OF SERVICE AGREEMENTS OF POPULAR ONLINE SERVICES

Even the shortest terms and conditions for popular online services are a few thousand words long.

As a result, **97%** of people in the 18-34 age group agree to conditions without reading them.

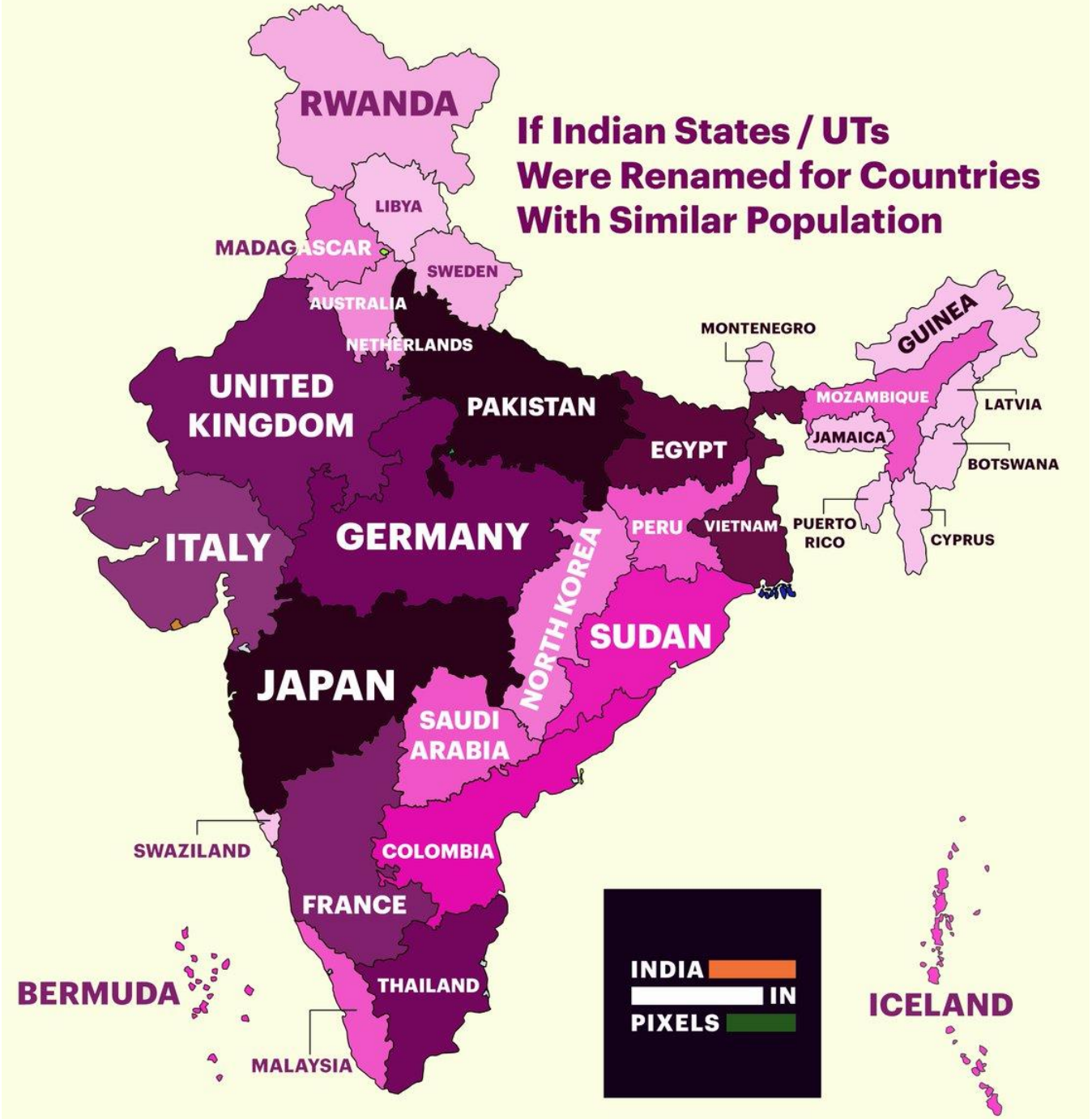
Reading time
Assuming 240 WPM

Length of text
Total word count

FACEBOOK INSTAGRAM SPOTIFY TWITTER LINKEDIN TINDER YOUTUBE APPLE AMAZON ZOOM TIKTOK NETFLIX MICROSOFT UBER

Assignment 3: Comparative Visualization

Transpose -- For impact & surprise, transposition data in unexpected ways.


If Indian States / UTs
Were Renamed for Countries
With Similar Population
RWANDA
LIBYA
MADAGASCAR
SWEDEN
AUSTRALIA
NETHERLANDS
UNITED KINGDOM
PAKISTAN
MONTENEGRO
GUINEA
MOZAMBIQUE
LATVIA
JAMAICA
EGYPT
BOTSWANA
ITALY
GERMANY
PERU
VIETNAM
PUERTO RICO
CYPRUS
NORTH KOREA
SUDAN
JAPAN
SAUDI ARABIA
SWAZILAND
COLOMBIA
FRANCE
THAILAND
BERMUDA
MALAYSIA
INDIA IN PIXELS
ICELAND

Assignment 3: Comparative Visualization

Translate -- Apply appropriate measurement scales to comprehend the data. Raw numbers to percentages, time into years, months, days, hours, minutes or seconds, lengths/distances into cms, mts or kms, areas into sq.units, volumes, weights, temperatures, and so on. Negative & log scales.

Transform -- Convert measurement scales into something more familiar & relatable to the reader.

Transpose -- For impact & surprise, transposition data in unexpected ways.

In this assignment, take a very small dataset and create 3 visualizations by applying translation, transformation and transposition of the data. Post your submission as a comment in this issue.

Data Sense

Approximately how many accidental deaths occur in the Mumbai suburban train network (Mumbai Local)?

a. 1200
b. 1800
c. 2500
d. 3500

Data Sense

What is Mumbai's approximate current population?

a. 18 million
b. 22 million
c. 24 million
d. 26 million

Data Sense

During peak hours, how many people pack into one compartment of a typical Mumbai local train?

a. 300
b. 400
c. 500
d. 600

Data Sense

During peak hours, how many people pack into one compartment of a typical Mumbai local train?

a. 300
b. 400
c. 500
d. 600

Do you know how much it is in terms of number of people per square meter?

Do you know the actual capacity of a compartment?

Data Sense

During peak hours, how many people pack into one compartment of a typical Mumbai local train?

a. 300
b. 400
c. 500
d. 600

Do you know how much it is in terms of number of people per square meter?
(14-16 per sq.mt)

Do you know the actual capacity of a compartment?
(188)

Data Sense

How many cases are pending in various judicial courts of Mumbai?

a. 3 lakhs
b. 6 lakhs
c. 12 lakhs
d. 24 lakhs

Data Sense

How many cases are pending in various judicial courts of Mumbai?

a. 3 lakhs
b. 6 lakhs
c. 12 lakhs
d. 24 lakhs

Do you know how of them are criminal cases?

Data Sense

How many cases are pending in various judicial courts of Mumbai?

a. 3 lakhs
b. 6 lakhs
c. 12 lakhs
d. 24 lakhs

Do you know how of them are criminal cases?
(40,000)

Data Sense

What is the car ownership in India?

a. 18 per 1000
b. 34 per 1000
c. 58 per 1000
d. 72 per 1000

Data Sense

What is the car ownership in India?

a. 18 per 1000
b. 34 per 1000
c. 58 per 1000
d. 72 per 1000

Do you know that this is lower than Sudan & Afghanistan

China? Europe? U.S?

Data Sense

What is the car ownership in India?

a. 18 per 1000
b. 34 per 1000
c. 58 per 1000
d. 72 per 1000

Do you know that this is lower than Sudan & Afghanistan

China? Europe? U.S?
(58, >500, >800)

Data Sense

What is number of tax payers in relation to voters?

a. 4%
b. 7%
c. 12%
d. 18%

Data Sense

How much did India's Mars programme cost?

a. $ 44 million
b. $ 54 million
c. $ 64 million
d. $ 74 million

Data Sense

How much did India's Mars programme cost?

a. $ 44 million
b. $ 54 million
c. $ 64 million
d. $ 74 million

NASA's Maven mission – $ 671 million
ESA Mars mission – $ 386 million
Japan's failed Mars mission – $ 186 million

Data Sense

How much did India's Mars programme cost?

a. $ 44 million
b. $ 54 million
c. $ 64 million
d. $ 74 million

NASA's Maven mission – $ 671 million
ESA Mars mission – $ 386 million
Japan's failed Mars mission – $ 186 million

Rs. 7 per km

Data Sense

How much did India's Mars programme cost?

a. $ 44 million
b. $ 54 million
c. $ 64 million
d. $ 74 million

NASA's Maven mission – $ 671 million
ESA Mars mission – $ 386 million
Japan's failed Mars mission – $ 186 million

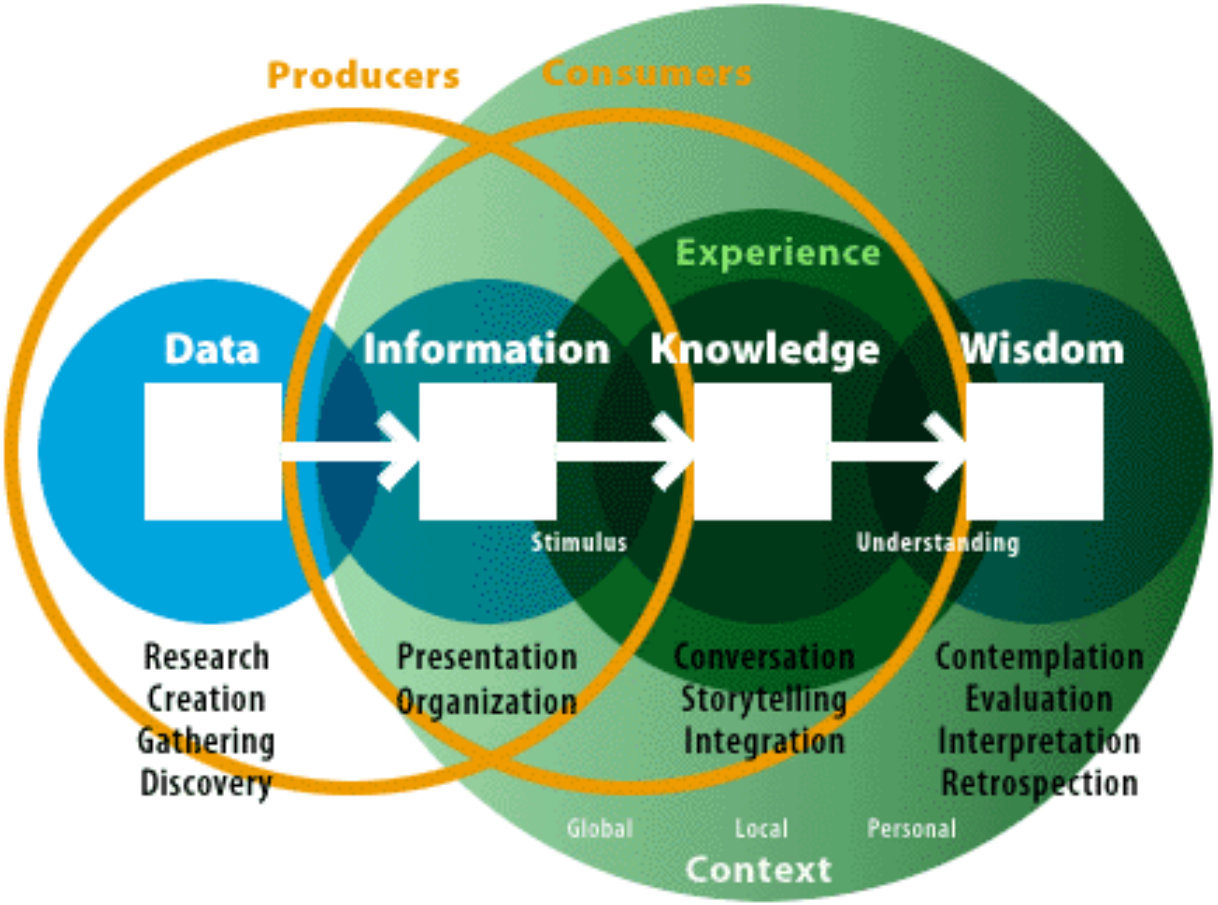
Rs. 7 per km (Mumbai auto Rs. 14.20 per km)

Data Sense

Ratio of fake doctors to real doctors practicing in Delhi:

a. 1:4
b. 1:3
c. 1:2
d. 1:1

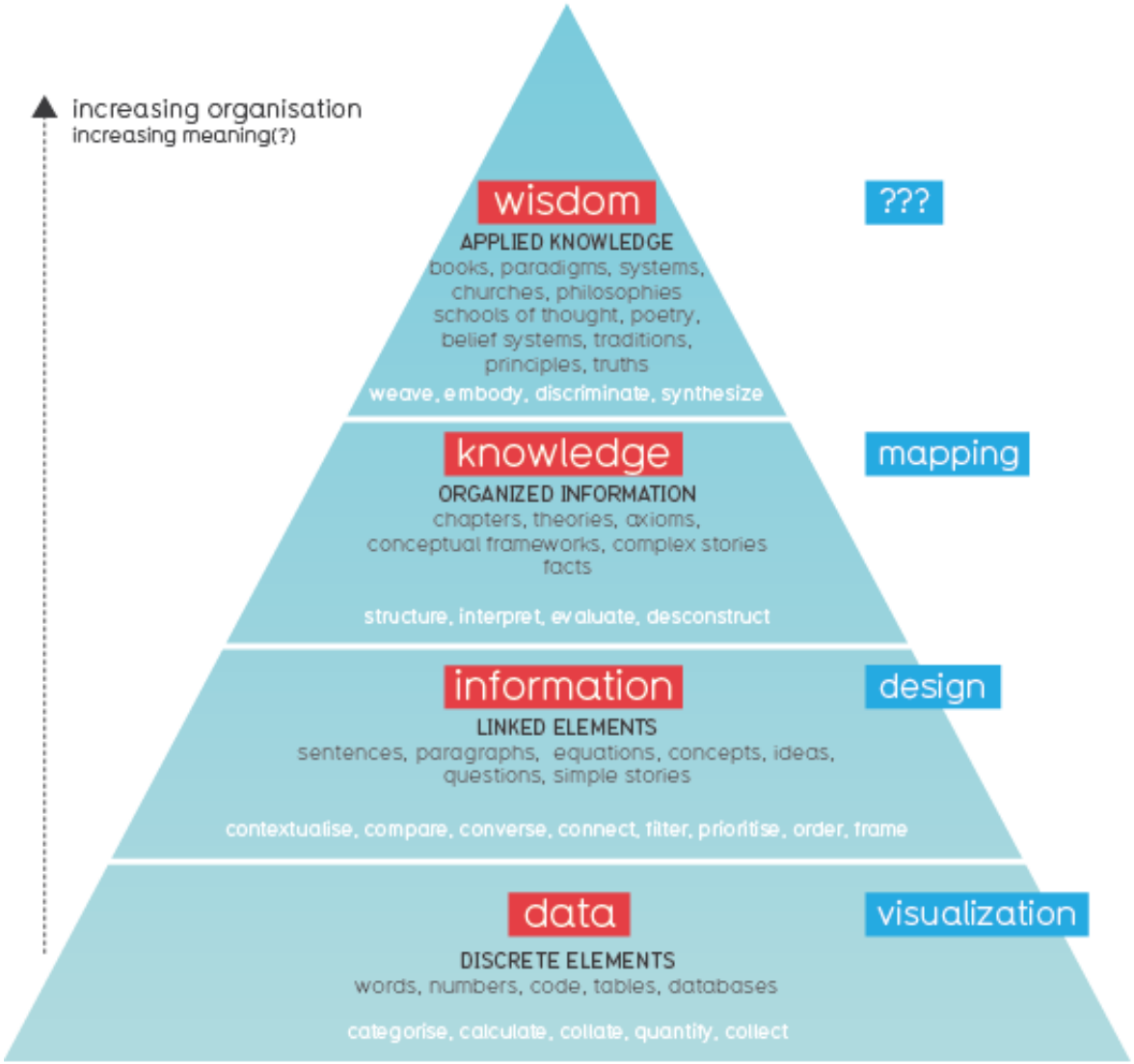
Intelligence Hierarchy: Data, Information, Knowledge, Wisdom



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 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

How Measurement led to the Modern World

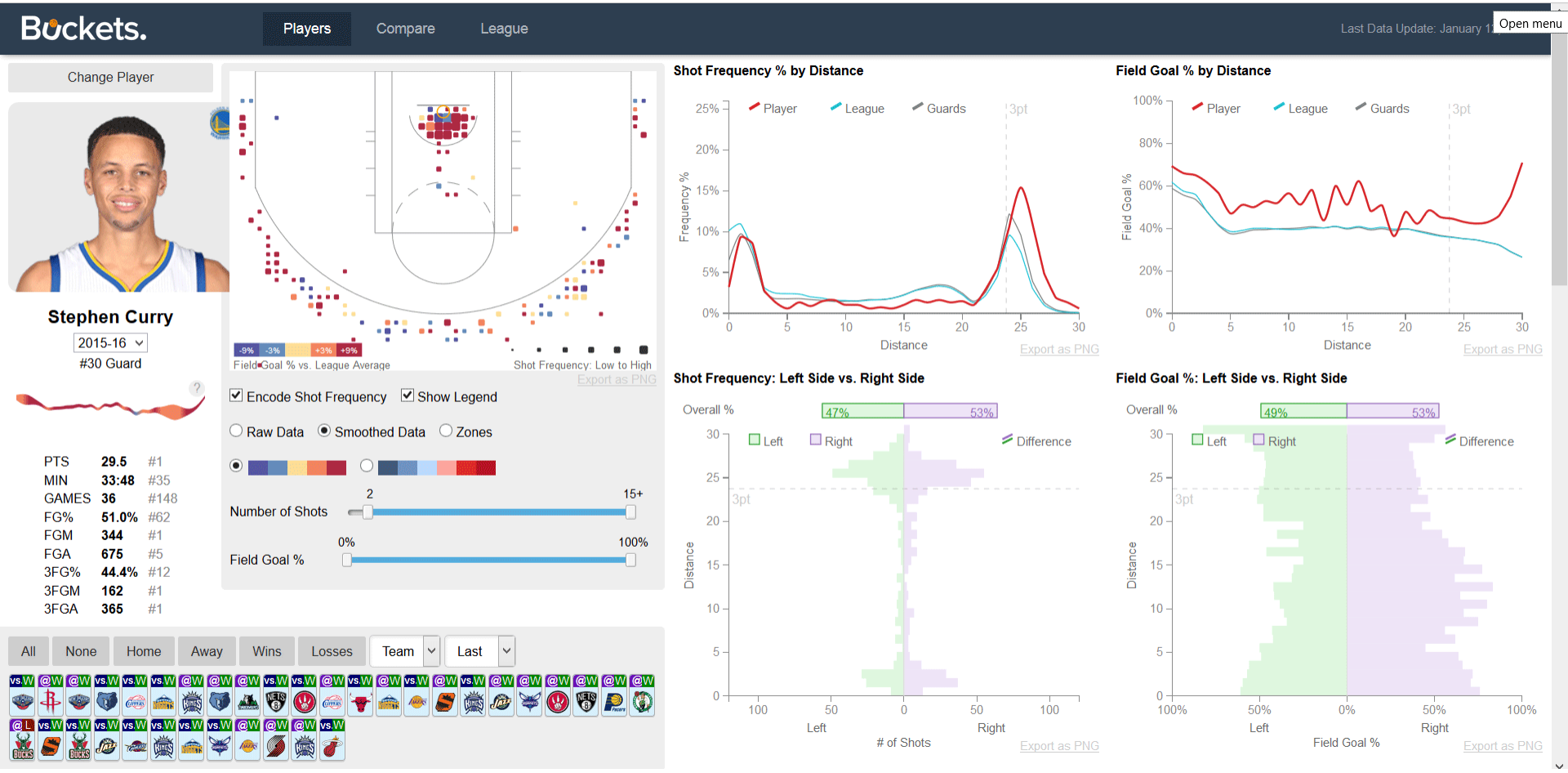
Pre-modern customs were all about dealing with trust, the need for direct supervision, and facing up to the enormous risks posed by nature.

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.

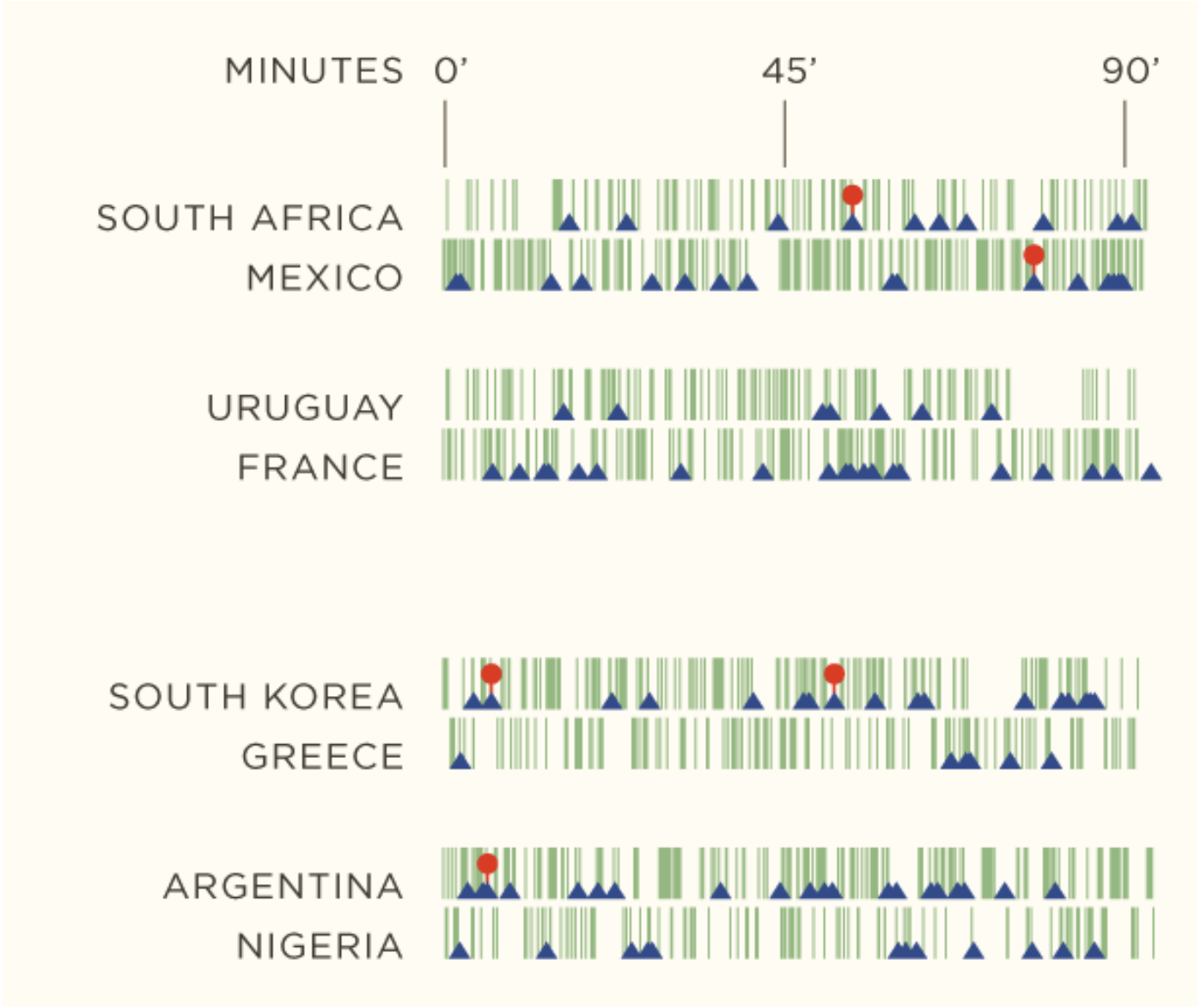
This ability to cheaply measure ushered in the world of modern institutions.

— 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)

Characteristics of data

Data Semantics are the real-world meanings of the data, such as whether the data represents temperature or height measurements. Focuses on organizing data that reflects the basic meaning, rather than solely or primarily on the relationships and attributes of the data.

Data Behaviour encompasses the trends, patterns, and shape of the data values relative to each other.

Data Type is the level or scale of measurement that describes the nature of information within the numbers assigned to variables.

Derived Data are created from the original data values. These include measures such as count, summation, or average, transformations such as binning, or the inclusion of additional, external data.

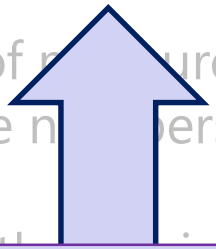
Reflections on how designers design with data. Alex Bigelow, Steven Druckery, Danyel Fishery, Miriah Meyer. AVI '14 Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, pp. 17-24.

Characteristics of data

Data Semantics are the real-world meanings of the data, such as whether the data represents temperature or height measurements. Focuses on organizing data that reflects the basic meaning, rather than solely or primarily on the relationships and attributes of the data.

Data Behaviour encompasses the trends, patterns, and shape of the data values relative to each other.

Data Type is the level or scale of measurement that describes the nature of information within the numbers assigned to variables.



The first two characteristics, semantics and behaviour, are inherent to data and are not decided or modified in the process of creating a visualization. They are, however, critical for a designer to understand.

Characteristics of data

The latter two characteristics are an important part of the design decision process when working with data. Specifically, designers have the freedom to create, define, and modify data structure and derived data.

Data Behaviour encompasses the trends, patterns, and shape of the data values relative to each other.

Data Type is the level or scale of measurement that describes the nature of information within the numbers assigned to variables.

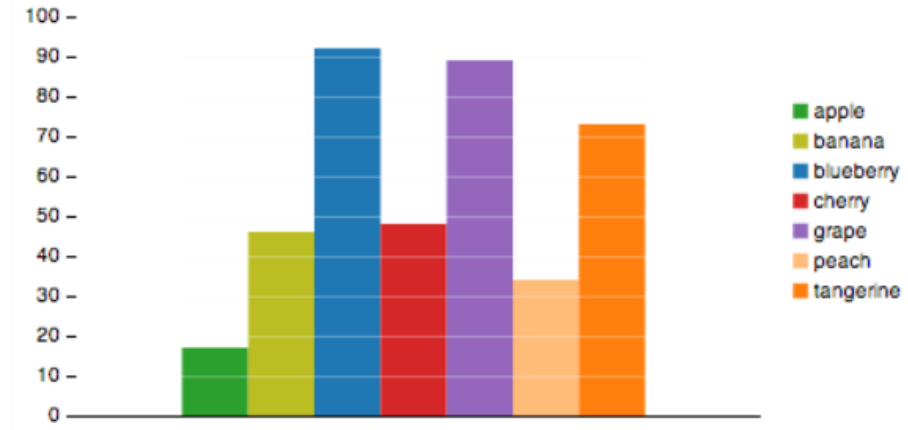
Derived Data are created from the original data values. These include measures such as count, summation, or average, transformations such as binning, or the inclusion of additional, external data.

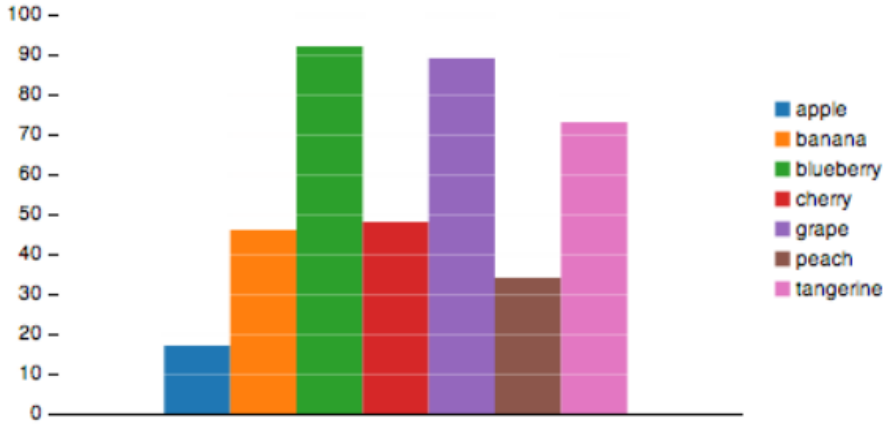
Characteristics of data

Data Semantics are the real-world meanings of the data, such as whether the data represents temperature or height measurements.

- Colour coding (semantically resonant colours)
- Data views (Vietnam War memorial)

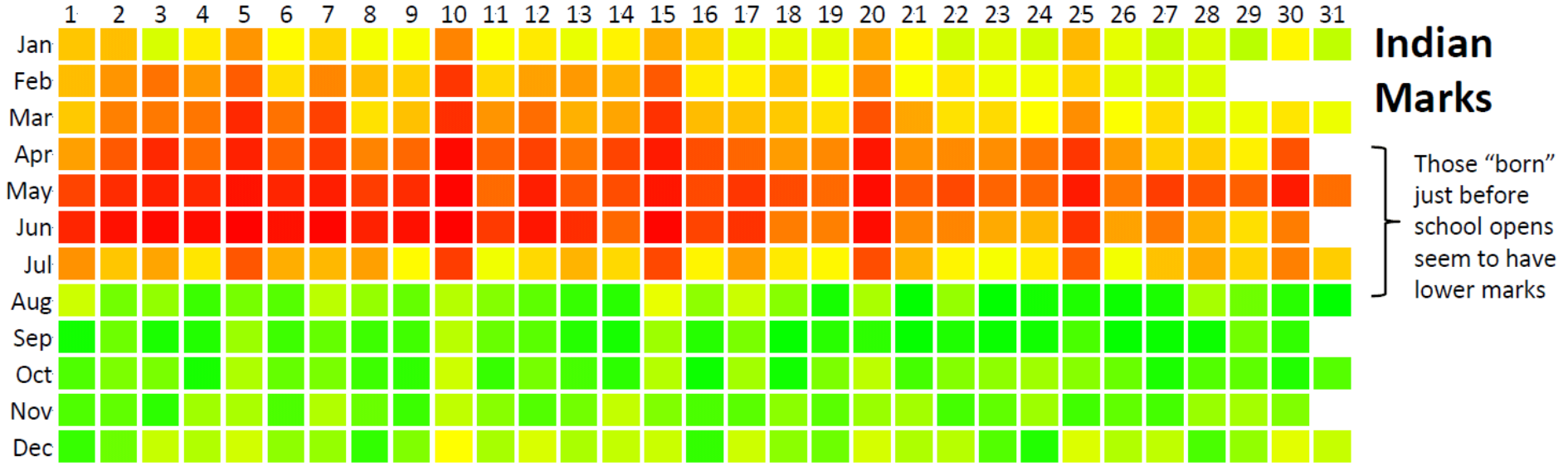
Data Semantics





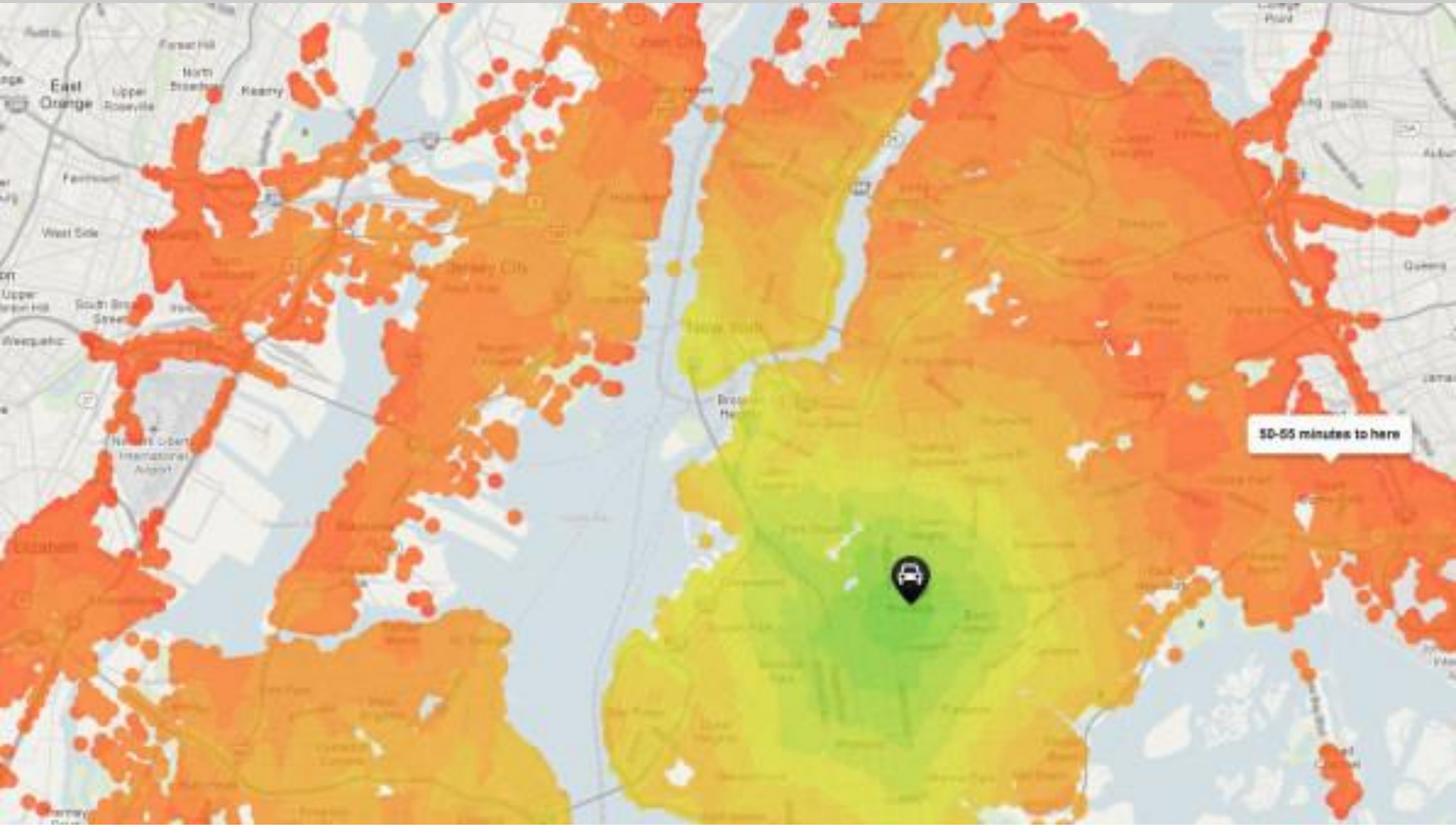
Lin, Sharon, et al. "Selecting Semantically-Resonant Colors for Data Visualization." Eurographics, vol. 32, 2013, pp. 401–410.

Data Semantics

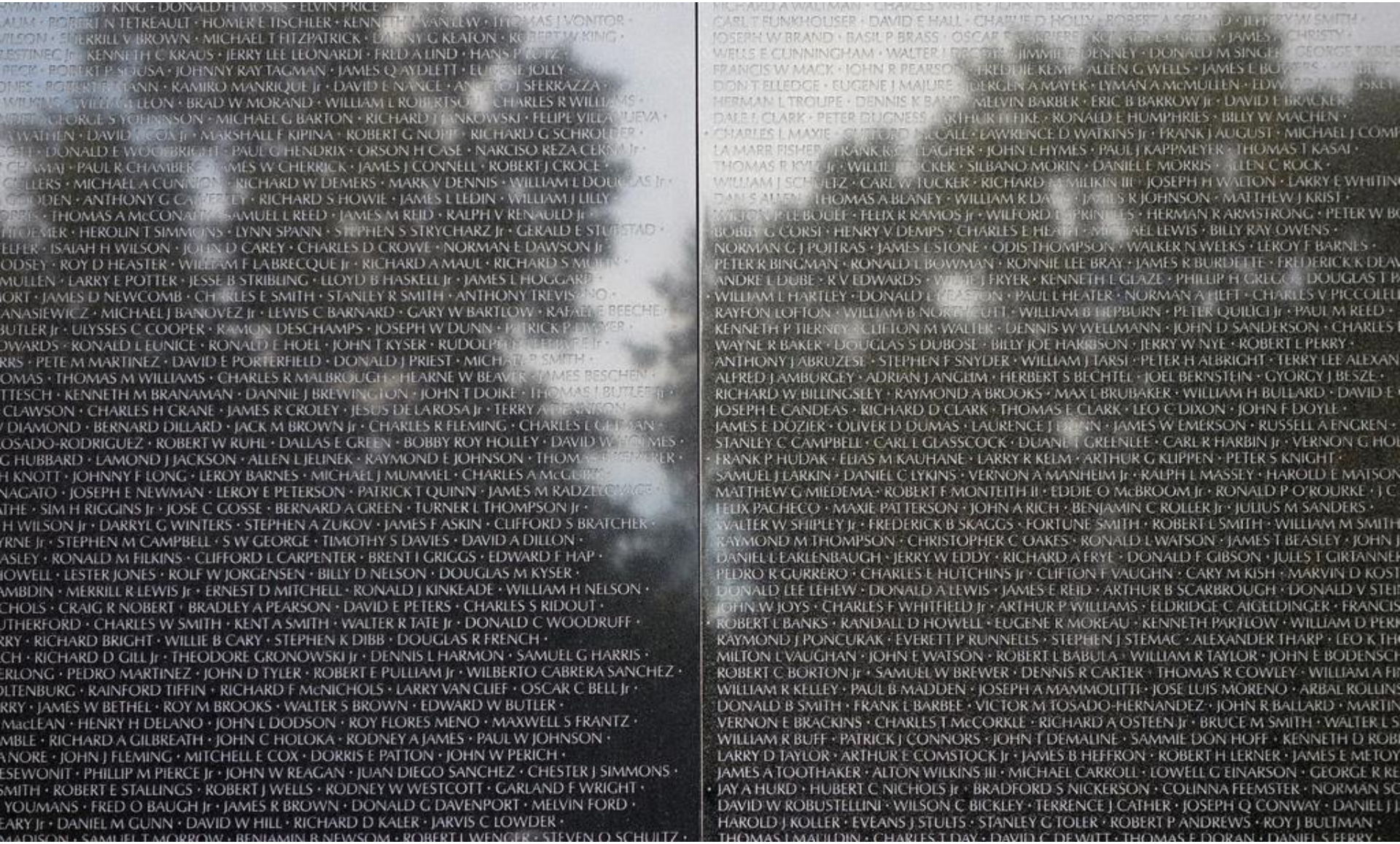


Birthdays in the U.S and in India, Gramener

Data Semantics



https://www.fastcompany.com/3062182/how-to-use-color-to-prove-your-point-from-a-data-viz-expert



Characteristics of data

Data Type is the level or scale of measurement that describes the nature of information within the numbers assigned to variables.

It refers to the interpretation of an item in terms of scales of its measurement (attributes)

Scales of 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.

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"				✔

Sheets | Charts | SmartArt Graphics | WordAr

	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Sheets | Charts | SmartArt Graphics | WordAr

	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	[illegible]	12/18/06	5-Low	Small Box	0.59	12/23/06
16	[illegible]	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	[illegible]	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Item

Sheets | Charts | SmartArt Graphics | WordAr

	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	[illegible]	12/18/06	5-Low	Small Box	0.59	12/23/06
16	[illegible]	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	[illegible]	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Item

Sheets | Charts | SmartArt Graphics | WordAr

	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	[illegible]	2/22/08
4	32	7/16/07	2-High	Small Pack	[illegible]	7/17/07
5	32	7/16/07	2-High	Jumbo Box	[illegible]	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	[illegible]	12/18/06	5-Low	Small Box	0.59	12/23/06
16	[illegible]	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	[illegible]	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Attribute

Item

Attribute Types

- Nominal
- Ordinal
- Interval
- Ratio

Attribute Types

- Nominal
- Ordinal
- Quantitative
  - Interval
  - Ratio

Attribute Types

N – Nominal/Categorical (labels)
apples, oranges, …

O – Ordinal
movie rating: 1 star to 5 stars

Q – Interval (location of zero arbitrary)
Jan 20, 2016; location: (LAT 47 LONG 122)
a geometric point
cannot compare directly, only differences (i.e. intervals)
may be compared

Q – Ratio (zero fixed)
length, mass, …
counts and amounts

Sheets | Charts | SmartArt Graphics | WordAr

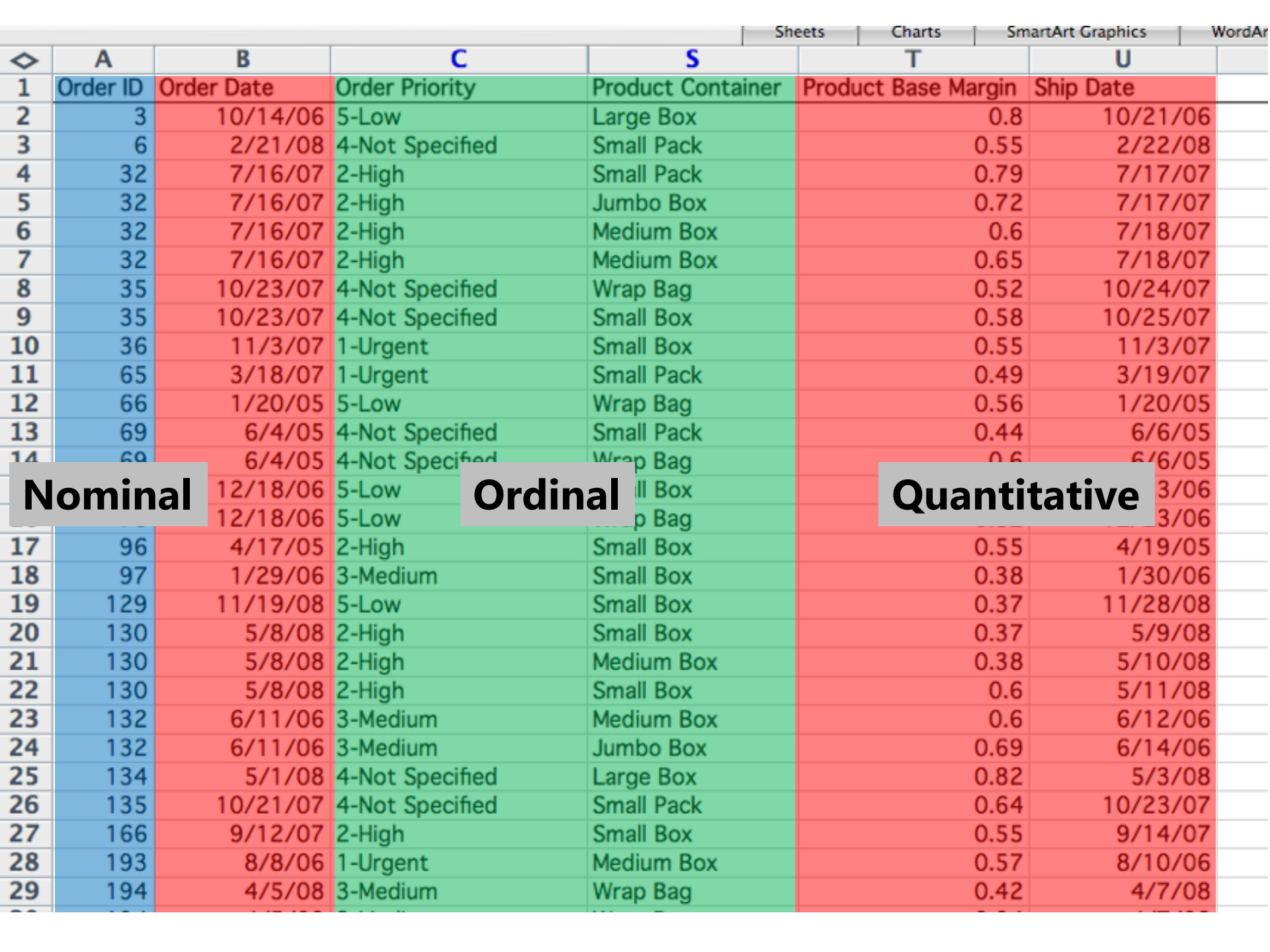
	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Sheets | Charts | SmartArt Graphics | WordAr

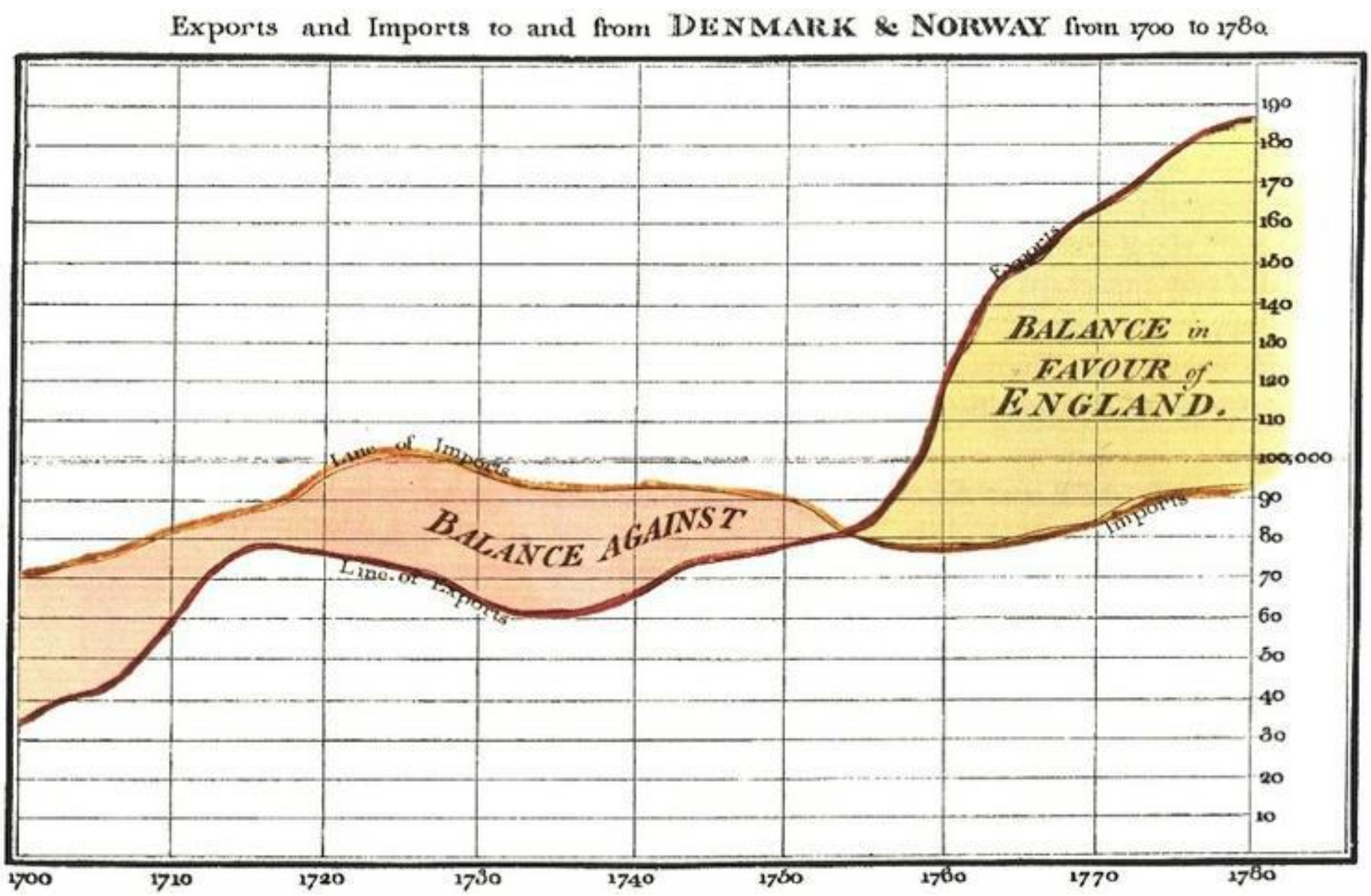
	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Sheets | Charts | SmartArt Graphics | WordAr

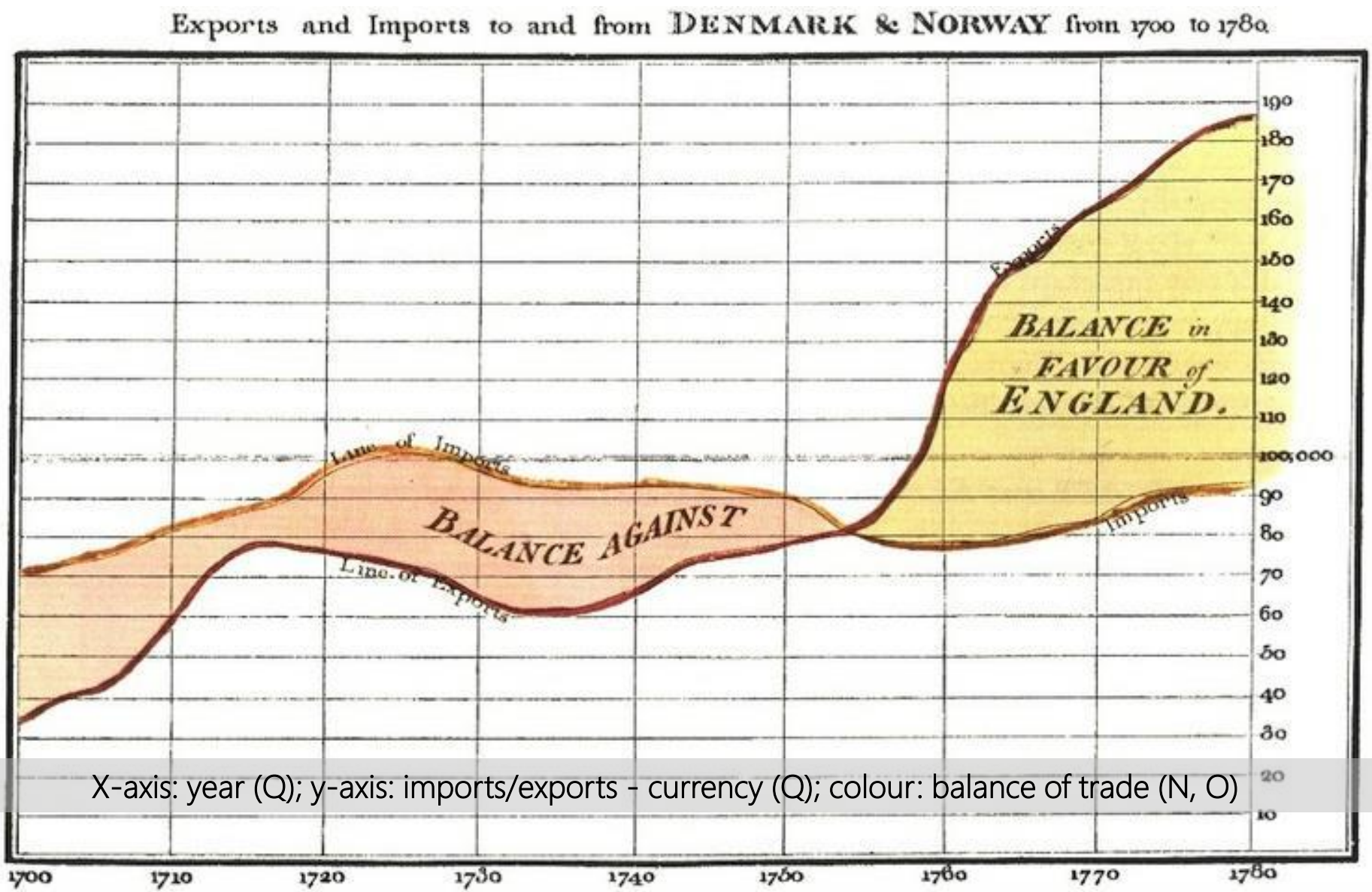
	A	B	C	S	T	U
1	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
2	3	10/14/06	5-Low	Large Box	0.8	10/21/06
3	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
4	32	7/16/07	2-High	Small Pack	0.79	7/17/07
5	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
6	32	7/16/07	2-High	Medium Box	0.6	7/18/07
7	32	7/16/07	2-High	Medium Box	0.65	7/18/07
8	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
9	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
10	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
11	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
12	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
13	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
14	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
15	70	12/18/06	5-Low	Small Box	0.59	12/23/06
16	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
17	96	4/17/05	2-High	Small Box	0.55	4/19/05
18	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
19	129	11/19/08	5-Low	Small Box	0.37	11/28/08
20	130	5/8/08	2-High	Small Box	0.37	5/9/08
21	130	5/8/08	2-High	Medium Box	0.38	5/10/08
22	130	5/8/08	2-High	Small Box	0.6	5/11/08
23	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
24	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
25	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
26	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
27	166	9/12/07	2-High	Small Box	0.55	9/14/07
28	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
29	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08


Sheets
Charts
SmartArt Graphics
WordAr
A
B
C
S
T
U
Order ID
Order Date
Order Priority
Product Container
Product Base Margin
Ship Date
3 10/14/06 5-Low Large Box 0.8 10/21/06
6 2/21/08 4-Not Specified Small Pack 0.55 2/22/08
32 7/16/07 2-High Small Pack 0.79 7/17/07
32 7/16/07 2-High Jumbo Box 0.72 7/17/07
32 7/16/07 2-High Medium Box 0.6 7/18/07
32 7/16/07 2-High Medium Box 0.65 7/18/07
35 10/23/07 4-Not Specified Wrap Bag 0.52 10/24/07
35 10/23/07 4-Not Specified Small Box 0.58 10/25/07
36 11/3/07 1-Urgent Small Box 0.55 11/3/07
65 3/18/07 1-Urgent Small Pack 0.49 3/19/07
66 1/20/05 5-Low Wrap Bag 0.56 1/20/05
69 6/4/05 4-Not Specified Small Pack 0.44 6/6/05
Nominal
Ordinal
Quantitative
96 4/17/05 2-High Small Box 0.55 4/19/05
97 1/29/06 3-Medium Small Box 0.38 1/30/06
129 11/19/08 5-Low Small Box 0.37 11/28/08
130 5/8/08 2-High Small Box 0.37 5/9/08
130 5/8/08 2-High Medium Box 0.38 5/10/08
130 5/8/08 2-High Small Box 0.6 5/11/08
132 6/11/06 3-Medium Medium Box 0.6 6/12/06
132 6/11/06 3-Medium Jumbo Box 0.69 6/14/06
134 5/1/08 4-Not Specified Large Box 0.82 5/3/08
135 10/21/07 4-Not Specified Small Pack 0.64 10/23/07
166 9/12/07 2-High Small Box 0.55 9/14/07
193 8/8/06 1-Urgent Medium Box 0.57 8/10/06
194 4/5/08 3-Medium Wrap Bag 0.42 4/7/08

Identify the data attributes

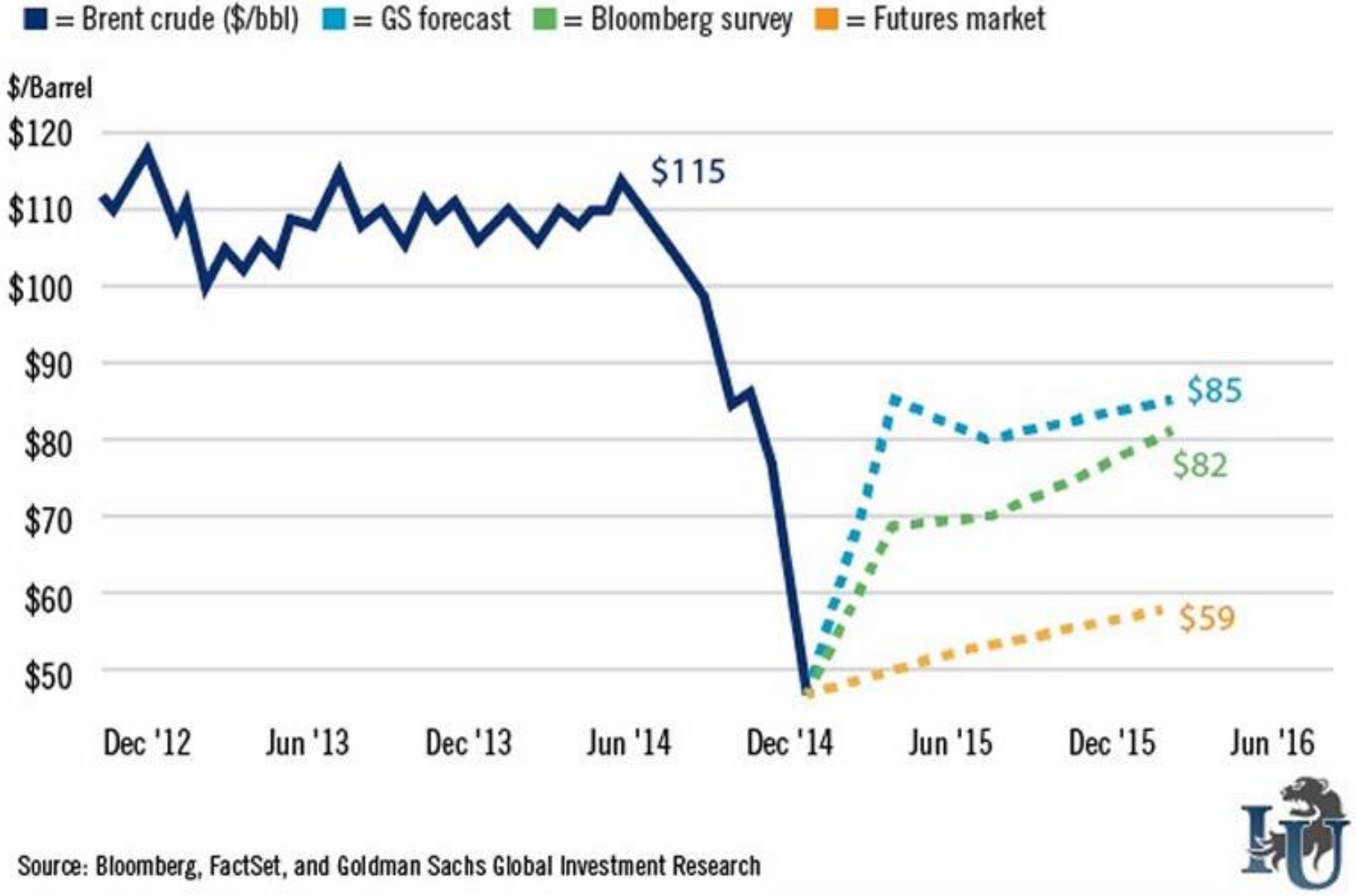


Identify the data attributes

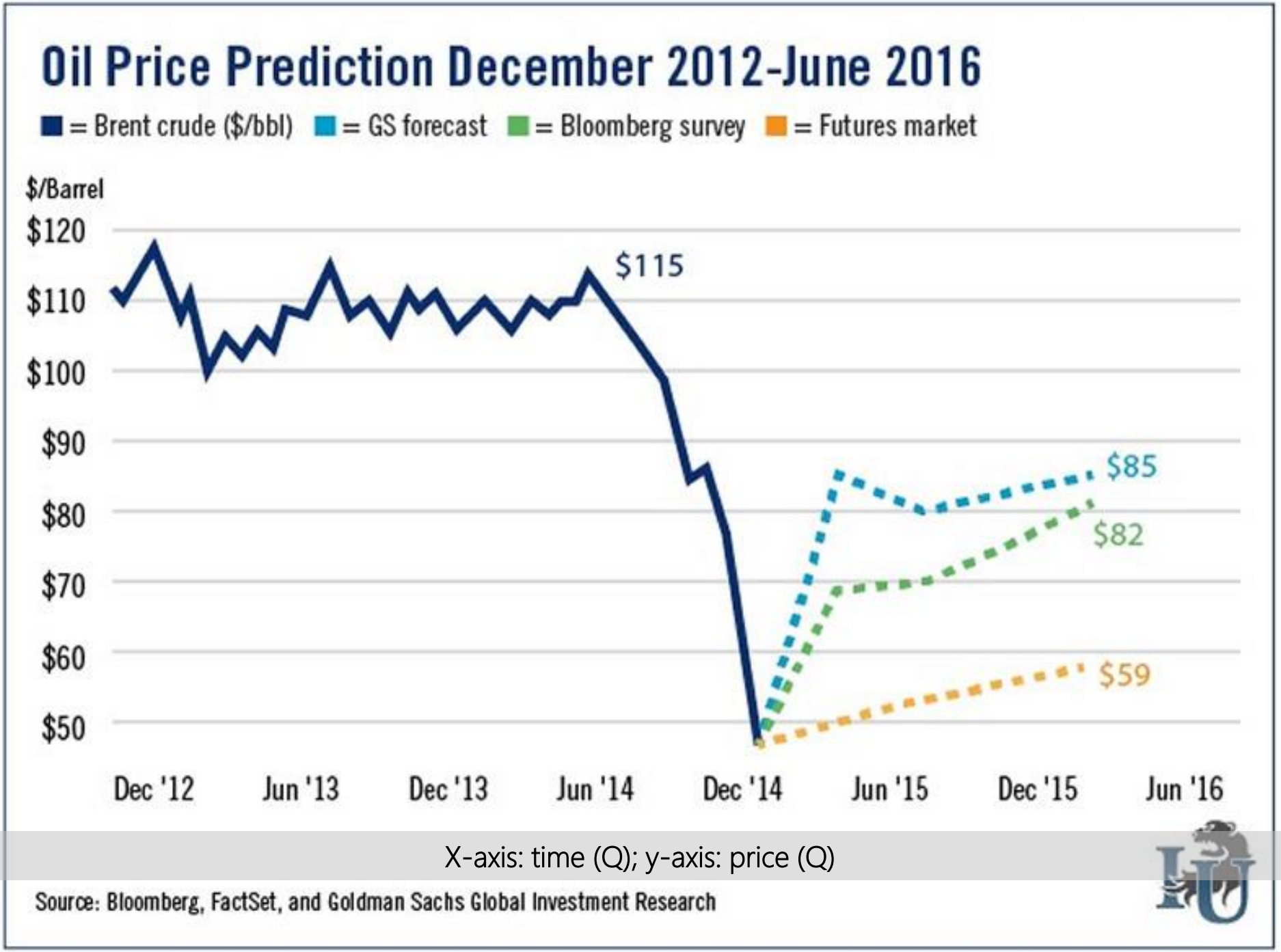


X-axis: year (Q); y-axis: imports/exports - currency (Q); colour: balance of trade (N, O)

Oil Price Prediction December 2012-June 2016



Source: Bloomberg, FactSet, and Goldman Sachs Global Investment Research



X-axis: time (Q); y-axis: price (Q)

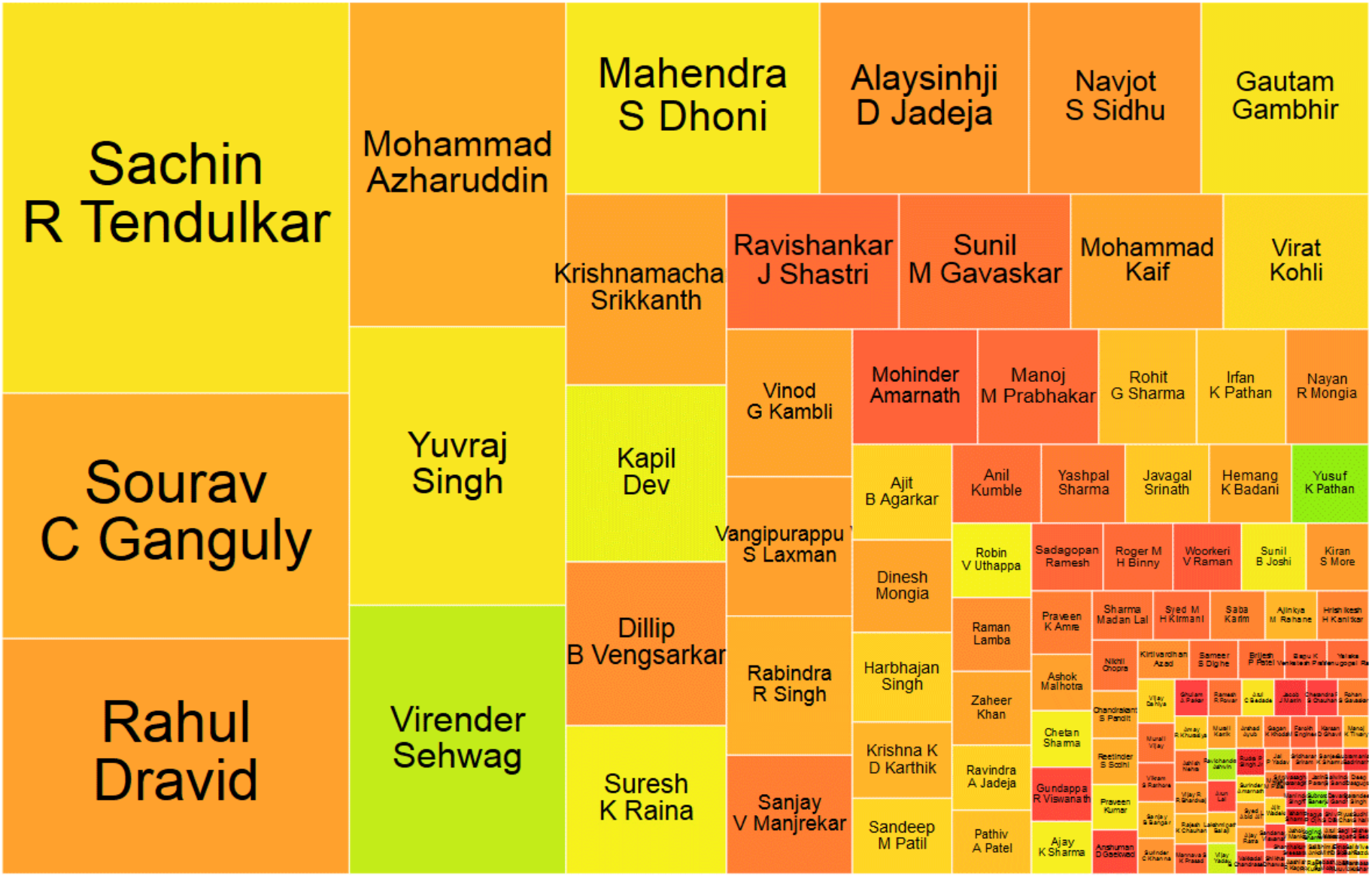
For the popular cricket-playing nations, we took every batsmen that has scored at least 20 runs in their ODI career, and plotted their run rate for every single match. Size = Number of runs. Colour = strike rate 20 50 80 110 140

ODI | Test

Ind | Pak | Aus | SL | WI | NZ | Eng | SA | Zim | World

Search

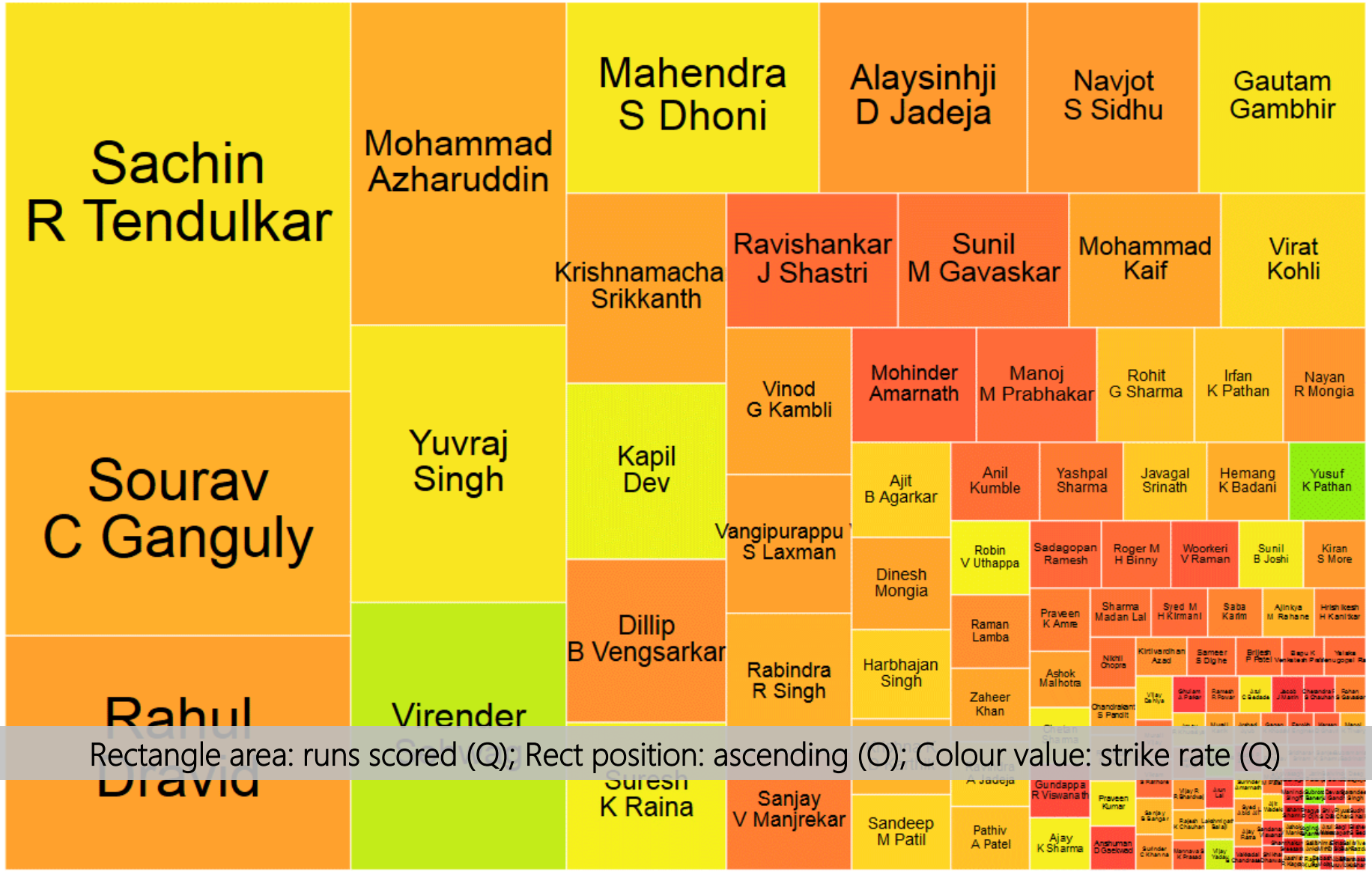
Expand | Collapse



For the popular cricket-playing nations, we took every batsmen that has scored at least 20 runs in their ODI career, and plotted their run rate for every single match. Size = Number of runs. Colour = strike rate 20 50 80 110 140

ODI Test | Ind Pak Aus SL WI NZ Eng SA Zim World

Search | Expand Collapse



Rectangle area: runs scored (Q); Rect position: ascending (O); Colour value: strike rate (Q)

Why is it necessary to understand data characteristics?

Why is it necessary to understand data characteristics?

1. In order to visualize data, data sets have to be mapped to visual attributes (also known as data encoding)

Why is it necessary to understand data characteristics?

1. In order to visualize data, data sets have to be mapped to visual attributes (also known as data encoding)

2. Understanding data types helps us determine which visual attributes represent which data types most effectively

Why is it necessary to understand data characteristics?

1. In order to visualize data, data sets have to be mapped to visual attributes (also known as data encoding)

2. Understanding data types helps us determine which visual attributes represent which data types most effectively

3. Classification of data with implications on the mathematical operations we can perform on them and how to structure them

Why is it necessary to understand data characteristics?

1. In order to visualize data, data sets have to be mapped to visual attributes (also known as data encoding)

2. Understanding data types helps us determine which visual attributes represent which data types most effectively

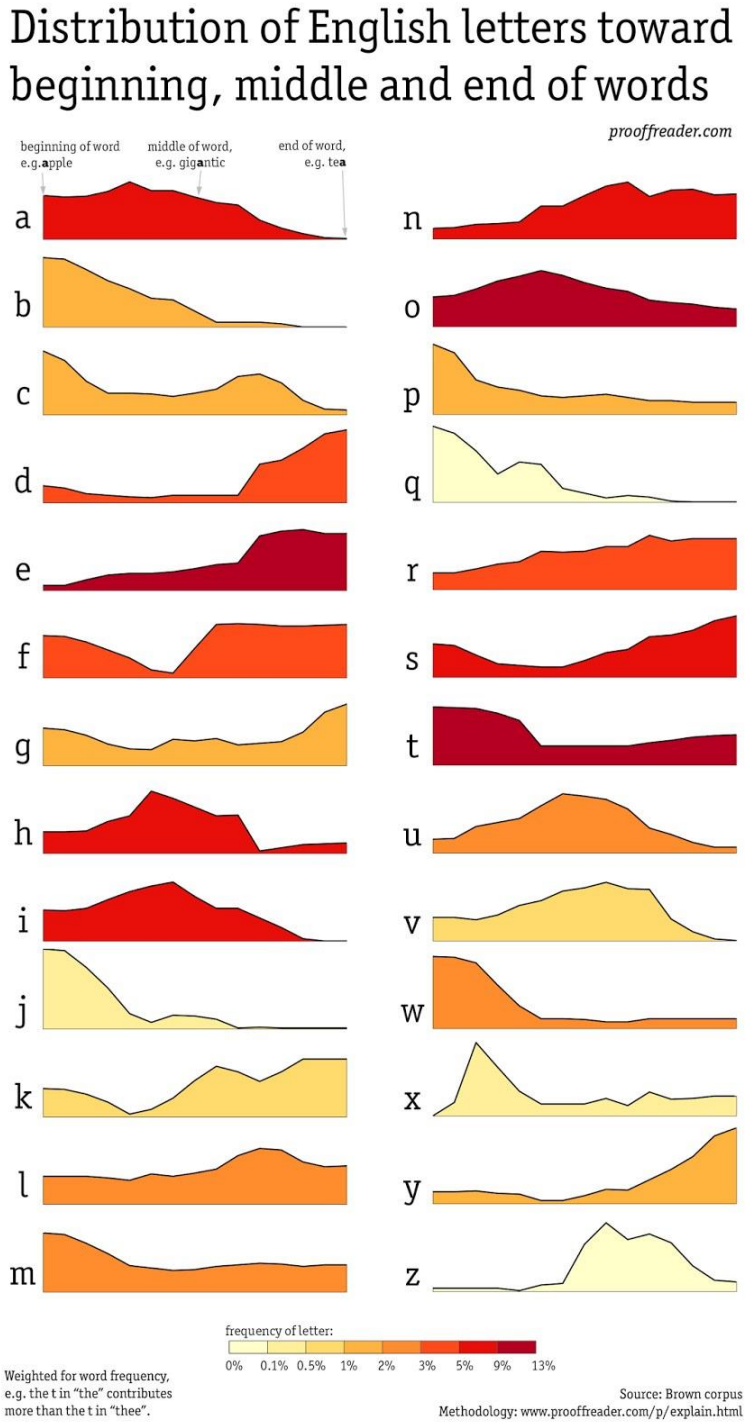
3. Classification of data with implications on the mathematical operations we can perform on them and how to structure them

4. Good data models result in good mental models, leading to better understanding and reasoning

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"				✔

David Taylor -- http://prooffreaderplus.blogspot.com/2014/05/methodology-and-analysis-of-letter.html

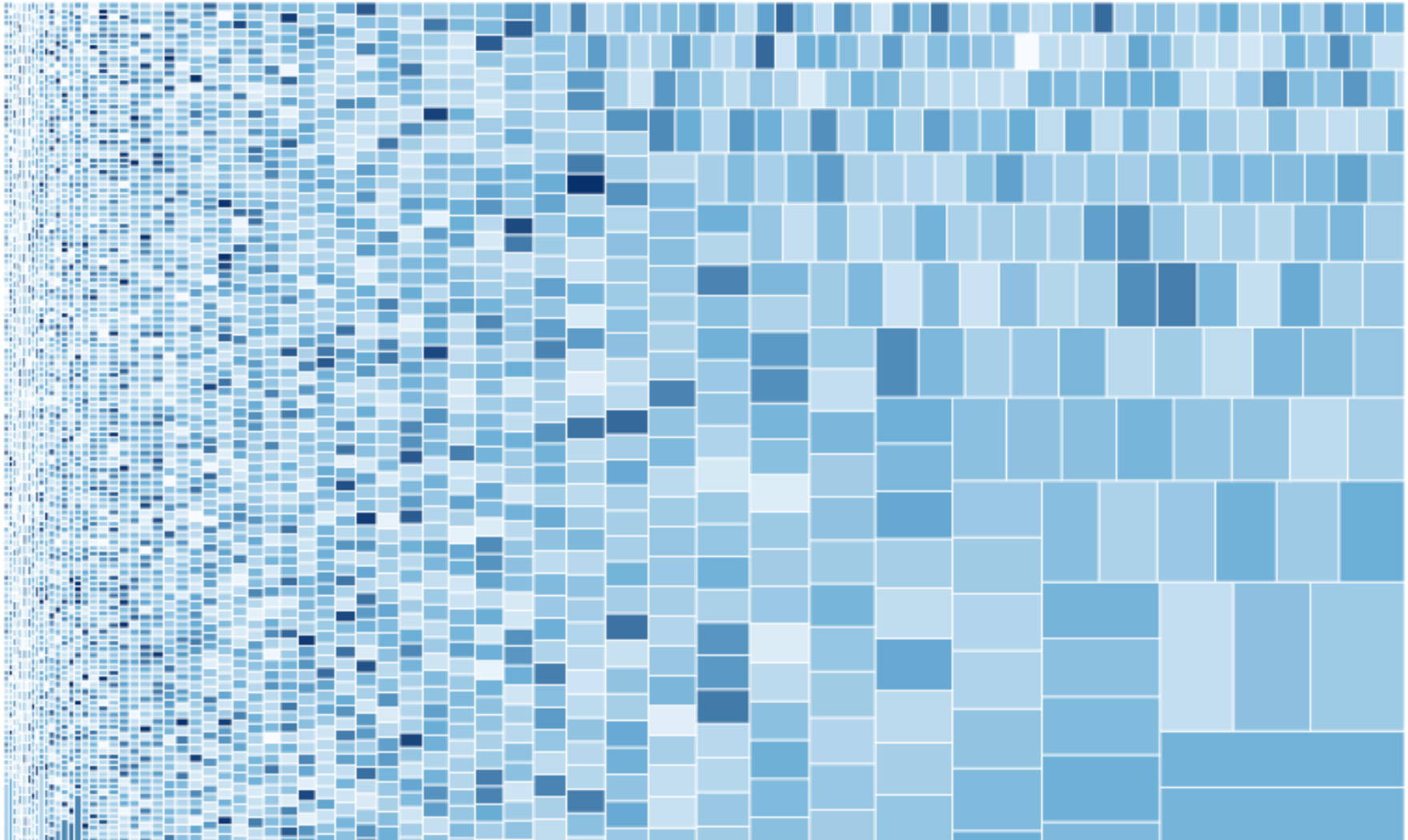


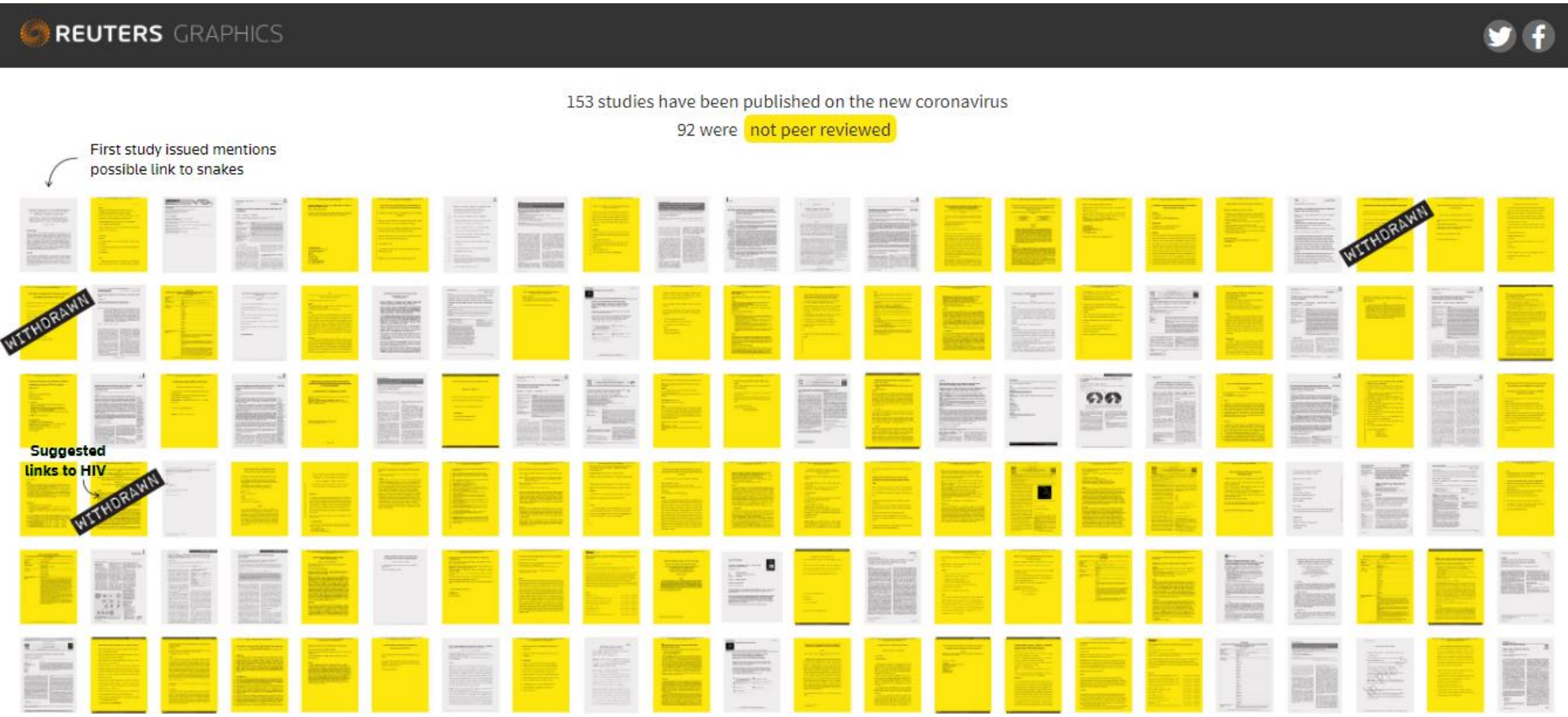
NAMES AND MARKS

Gramener

Here are the top 5,000 names of students that appeared in the Class XII State Board examinations in Tamil Nadu. The larger boxes show popular names. The colour of the box indicates the average percentage scored by the student in the Board exams.

Search for a name | Low marks | High marks | Clear





https://graphics.reuters.com/CHINA-HEALTH-RESEARCH/0100B5ES3MG/index.html

Volume I: Russia interference in the 2016 presidential election

Volume II: Obstruction of justice inquiry

https://www.nytimes.com/interactive/2019/04/19/us/politics/redacted-mueller-report.html