Liu Yikun Dong Zhao

Visual Storytelling

Infographic Design in News

images
Publishing

CONTENTS

Visual Storytelling
Infographic Design in News

We are now in the age of New Media, in which journalism's role and influence are being transformed, and digital news agencies and digital reporting are booming. Mobile platforms and social media are changing the way news content is produced and disseminated, and one of the hottest new developments is the arrival of Big Data.

With the diversification of communication channels, the development of citizen journalism is playing an increasingly important role in affecting the timeliness of news delivery. The influence of traditional print media and broadcast news has declined, and traditional news coverage often falls short of the expectations of mass audiences for increasingly diverse information. Data journalism is helping to bridge this gap.

Data journalism is a new kind of news storytelling enabled by the vast amounts of digital information collected by companies and governments. Using such data, relationships—between people, organizations, and governments—can be quantified and visualized.

Data journalism contains two kinds of innovation. The first of these is the use of graphics and the opportunity for the reader to interact with the information they contain. The other innovation concerns content. Data journalism collects, filters, and presents information in new ways, discovering new, more meaningful patterns and relationships. A small amount of information in an isolated event often lacks obvious meaning, but if viewed in larger clusters, from new angles, data can reveal more important significance. Diverse presentation techniques of data journalism not only enrich the types of news coverage, but make the presentation more concise and intuitive. Data journalism has a far-reaching impact on news production, since news collecting and editing is not limited to a single form, but needs updated skills, including data journalism design, so as to adapt to the changing times.

Data journalism is the future of journalism. With improvements in data analysis, information extraction and data mining technology, journalism is better placed to add depth to the issues it covers and to deliver the news in more creative ways. This innovative dimension is mainly reflected in three areas—investigative journalism, data visualizing narration, and data-driven application. From everyday social issues to deep social trends, it helps to provide reliable insights and predictions. News presented in graphic form aids comprehension and can help create an emotional response from the reader.

A Origin and Characteristics of Data Journalism

Origin of News Data

According to the Pew Research Center's Internet and American Life Project, more than 90% of the world's news data in recent years was generated by government, institutions, enterprises (including the media), and Internet users. The content created includes text, images, audio and video, and location information. The data has four notable features—velocity, variety, volume, and value.

Characteristics of Data Journalism

1. Data journalism enables traditional news stories to be told using digital information. Specialist data journalists can help traditional journalists tell a complex story through engaging infographics; and open paths to new kinds of stories.

2. Data journalism can help explain how a story relates to the individual, as the BBC and the *Financial Times* now routinely do with their interactive budget reports (in which readers can find out how a particular budget affects them). It can also open up the news-gathering process itself, as *The Guardian* does so successfully in sharing data, context, and questions in its Data blog. News today is increasingly related to data and numbers, and the media's role is to report and explain it.

3. New journalism enables a journalist to report stories that have not previously been reported and to report them from new angles. This is done by finding and visualizing large amounts of data, thus reporting stories from new perspectives. Data journalism requires deep data mining through repeated extraction, screening, and reorganization, focusing on specific information to filter the data, and interpreting the information in order to visualize it and produce news stories.

4. The core of data journalism is data processing, which tends to have the following characteristics: it aims to serve the public interest; is based on open data; relies on special software for data processing to find news stories hidden in abstract data; and presents news in visual and interactive ways.

B Forms of Data Journalism

There are three main forms of data-based news: data visualization, graphic-based news, and data maps.

1. Data visualization is a very broad concept, referring mainly to graphical methods to clearly and effectively distribute and communicate information.

2. There is no clear boundary between graphic-based news and data visualization, because graphic news is data visualization in the broadest sense. The difference between them is that news graphics simply illustrate the essential details of the traditional news story, while data visualization always focuses on statistics and presentation of digital information.

3. Data maps usually integrate a variety of information but are mainly used in the reporting of disasters such as earthquakes, tsunamis, and landslides.

The media tend to report data-based news visually because it is a more effective way to communicate such information. Taking advantage of computer technology, data journalism visualization differs from the style of graphics traditionally used in news reporting. To these innovative new forms of graphical presentation the media have added multimedia, which combines video, text, and interactive graphics. These forms and delivery platforms not only encourage readers to engage with the information, they perfectly suit the fast-paced, multi-platform nature of the Internet.

Data journalism has transformed the way traditional journalism is done and consists of four steps: digging deep into data, filtering the data to find specific information, visualizing the data, and writing the story.

C Data Journalism in the Media: Case Studies

The Guardian

The Guardian is a pioneer in introducing and developing data journalism, and has had a big impact on other successful global news media, including the BBC, *The New York Times*, *Bloomberg News*, *The Wall Street Journal*, *Financial Times*, *The Associated Press*, *Texas Tribune* and *Mother Jones*, which have all launched their own data journalism columns.

The Guardian has produced thousands of pieces of data news, covering politics, the economy, sports, war, disaster, the environment, culture, fashion, science, technology, health and other fields. These take the form of charts, maps, and a variety of interactive renderings. Data news generally has the following features:

1. Participation

The newspaper opens the data platform and invites users to participate in data news coverage. *The Guardian* has been aware of the need for new forms of news delivery (namely 'open news') for a long time. Alan Rusbridger, editor-in-chief for 20 years, defines 'open media' as news that is fully woven into existing information networks in the world. It is connected to this network; it screens and filters data from the network; works closely with the network; and better explains the world using materials published and shared by ordinary people.

2. An Open Technology Platform

The Guardian invites third-party developers to participate in the value-added development of data, offering large amounts of content free of charge for their commercial use. This includes millions of articles,

photographs, and videos created since 1999, as well as a lot of public database information.

3. Enhance Advertising Effectiveness
The Guardian signed a cooperation agreement with the emerging targeting network advertising company Quantcast, planning to subdivide advertising audiences with the help of big data mining and analysis techniques. The goal was to assist advertisers to target their advertisements more accurately, improving their return on investment.

The New York Times

In April 2014 *The New York Times* launched a new column, The Upshot, which uses data analysis and presentation to help readers understand the meaning behind complex political and economic events. As well as serious political topics, it also introduced a number of sports-themed works, covering the World Cup, football, baseball, soccer, and the NBA. From these works, we can see the following features:

1. Individualization
Individual readers can engage with a piece of data news through interaction. For example, the article 'Is it better to rent or buy?' relates macro data such as the rise and fall of house prices to the life of every reader through interaction. This feature is also available in other sections of *The New York Times* as well as many other news organizations.

2. Functionality
Ideally, data visualization pieces should achieve the harmonious balance of aesthetics, utility, and news value, which is rarely achieved in general articles due to limited time and manpower. *The New York Times* uses concise visualization design to highlight the functionality of its pieces to readers and users.

3. Reorganization
Although not every media agency can organize a data journalism team of more than a dozen people like *The New York Times*, data journalism has undoubtedly become a key development area of many media agencies. In 2014, data journalism pioneer *The Guardian* reorganized its editorial team and announced data journalism as one of its three development directions. Newly created media organizations, such as Quartz, Vox, and Buzzfeed, are also building their data journalism teams.

Bloomberg News

Bloomberg's 'World's Billionaires' is a data news application based on network interaction to show the wealth of the world's top 100 billionaires. It combines the daily stock market ups and downs, macroeconomic changes, and Bloomberg's original reports to calculate the personal wealth of these billionaires, whose rankings and changes are constantly updated online. Visualization is presented in four ways: pictures, graphics, scatters, and maps. Full-time data journalists keep the information up-to-date. This application won a number of design and journalism awards, and two trends are very obvious:

1. Data-driven
Background data updating is the vital ingredient of a long-term data journalism project, creating more space for project development and encouraging more users to have their own ideas through the observation of data in news clues. Compared with the Forbes World's Billionaires list, which is updated once a year, Bloomberg's list may have surprising changes every single day, increasing the user retention rate.

2. Interaction
The application provides complete data in many different areas, providing users sufficient space to explore and discover. Readers can browse the information of the billionaires they are interested in through different indexes, such as profession, nationality, sex, age, and source of wealth.

Data Journalism Design

Basic Elements of Data Journalism Design

Different data presentation methods have different interpretative results, and data selection and application can directly influence the audience's understanding of the news. The basic elements of data journalism design are numbers, graphics, and colors.

1. Numbers
Around 1986, renowned psychologists Daniel Kahneman and Amos Tversky conducted a famous series of experiments to illustrate that different data presentation methods would affect the audience's choice behavior. The following are their questions and strategy options:

Imagine that the United States is facing an unusual epidemic, and 600 people are estimated to be dead. At present, there are two strategies, and it is assumed that scientists have precisely evaluated the results of each strategy. If strategy A is taken, 200 people will be saved; if strategy B is taken, it has a one-third chance of saving all the people and has a two-thirds chance of saving no people. The result is that 72% of the 600 people surveyed chose strategy A and 28% chose strategy B.

Another approach proved that if the results were the same but the data was described in different scenarios, participants would tend to choose the option that highlighted the people saved rather than the people who would die. If strategy A is taken, 400 people would die (22% of the participants chose this option); if strategy B is taken, it has a one-third chance of saving all the people and a two-thirds chance of saving no people (78% of the participants chose this option).

Thus, it can be seen that the way numbers are presented can directly affect readers' judgment on the central idea of the article. Data is the

001
Smoking Infographic

002
The Rise of the Silver Surfers

most important element in data journalism, and each step—from data collection to data presentation and finally to data description—may affect readers' understanding of the content.

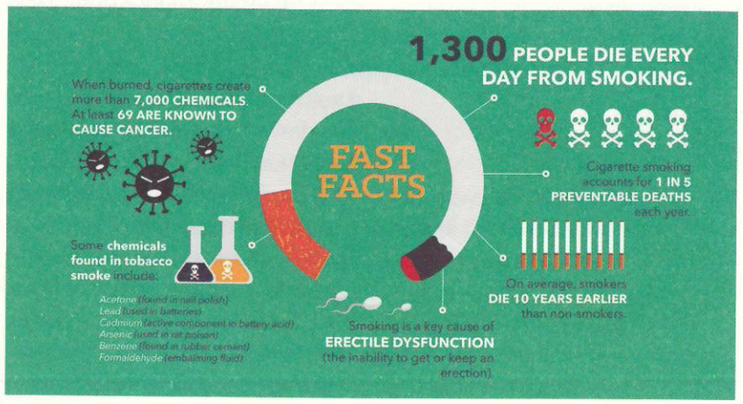
News media often uses digital strategies to set the media agenda. For example, when reporting the data of the International Monetary Fund (IMF), Peruvian newspaper *El Comercio* pointed out that China's purchasing power would be $17.6 trillion, beating the $16.4 trillion of the United States. This figure was calculated using IMF's latest purchasing power calculation method. However, if calculated by the original method, China still lagged $6.5 trillion behind the United States. This kind of data presentation reveals the framing effect of news media. If the newspaper wants to highlight China's rapid economic growth, they can use the latest IMF data, and if they want to express that China will not overtake the United States within a short period of time, they can use the unadjusted data.

2. Graphs

In addition to the presentation of numbers, the visual presentation of graphs can directly affect the audience's psychological reaction. When making infographics, rational use of geometric elements can increase the appeal of data, enriching the data with emotional color. The infographic in Figure 001 is designed to draw the reader's attention to the fact that smoking is harmful to health. Graphic symbols associated with death, such as 'skulls' and 'chemicals', were used to illustrate the statistics surrounding the health hazards caused by smoking. These graphic elements were adopted to encourage more vigilant behavior from readers when looking after their own health.

The graphs of data journalism often need other graphic elements to beautify pages and often use symbols rather than text to make the presentation more attractive. Figure 002 depicts the evolution of media used at different times by older people of 50 years old and above. In this graph, simple symbols are used instead of text classification to make the content clear at a glance and to make the news more attractive.

001



3. Color

Color selection for data journalism is also very important. Media around the world will turn their websites into black and white after catastrophes to express sympathy. During the National Day of China, major news websites adopt a red theme for celebration. The environmental protection theme is reflected in infographics by the use of green as the base color.

Though color has the least amount of impact among the three influencing factors on data presentation, its role should not be discounted. Different cultures associate different meanings to colors. For example, gray white can be used to represent funeral ceremonies in China, but represents purity and relates to marriage in Christian countries. Meanwhile, rainbow color is a symbol of homosexuality in the United States and other western countries. Today, the world is closely connected through the Internet, and news reports are no longer confined to domestic audiences. Therefore, there's a need for careful color selection in news reports involving international issues (particularly when dealing with topics such as race and gender).

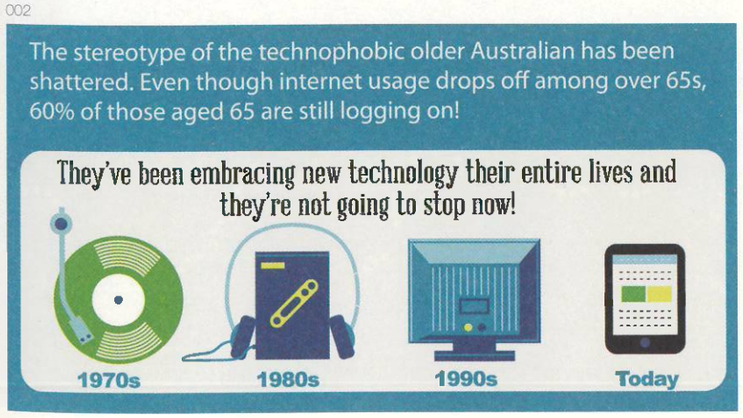
4. Other Related Concepts

Before we discuss the design and production of data journalism, it is necessary to clarify some concepts that are commonly used in statistics and scientific research methods, and are in line with international norms.

Data: Data is the information carrier—the numbers and characters describing the properties of things and a set of all the symbols that can be used for statistical analysis.

Index / variable: An index is a concept that explains overall quantitative features. An index can define the scope of a set of data. For example, Figure 003 shows the 'area' distribution of sugarcane yield, and the 'area' is an index that can be used to determine the interval of total amount of data. An index is also called variable in the data with mathematical significance.

002



003

Sugarland Distress

Project Description /
The graphic shows the quantity and growth of the sugar industry in India. Uttar Pradesh is India's largest state involved in the production of sugarcane, but it was in crisis at the time of publication. Santosh Kushwaha explored and illustrated the data by using physical pictures and some graphic elements to make it easily understood.

Client / *Business Today*, *India Today* Group
Design Agency / Visual Best
Designer / Santosh Kushwaha
Completion / 2013

004

Space Shuttle Missions Timeline

Project Description /
This interactive timeline is about NASA's space shuttle missions from 1977 to 2010. All the content is very clear and easy to navigate in this innovative visualization.

Client / lainformacion.com
Designer / Carlos Gámez Kindelán
Completion / 2014

Sample / total sample: The sample is a subset of observations taken from a parent population according to a random principle, and the size of the sample is the total sample. For example, India has 16 states and one capital district. If five main sugarcane production areas are taken to represent India's overall sugarcane yield, the sugarcane yield of these five areas is the statistical sample, and the sample size is five.

Case: Each of the individuals making up the sample is called a case.

Data value: Data value describes the sample data in numbers.

For example, Figure 003 shows the sugarcane yield all over India, and 'sugarcane yield' can be a set of data composed of the sugarcane yield of five areas in India. In this set of data, the total sample is the sugarcane sample of these five areas, the sugarcane yield of each area is one 'case', the area is one 'index', and the yield is 'data value' rather than 'data'.

Traditional Design Methods

Descriptive data is the most fundamental type of data journalism. Descriptive statistics is describing and summarizing the overall distribution of data using a variety of design methods, such as a line graph, radar chart, histogram, bar chart, pie chart, and scatter plot.

Figure 003 describes the development and current crisis of the sugar industry in India. This is a piece of descriptive-style news and it adopts a variety of design methods, including a line graph, pie chart, histogram, and bar chart.

Before working on the graphic design of the descriptive data, the first thing to consider is the integrity of the data. Data integrity means that the data can represent a complete concept within a certain range. For example, Figure 003 presents a picture of India's sugarcane economy and so depicts the sugarcane yield data of the whole country. In addition to the yield data of four main sugarcane-producing states—Bihar, Uttar Pradesh, Tamil Nadu, and Karnataka—the author included information on states with smaller sugarcane yields, but used 'Others' to present the overall yield situation, so that the data is complete.

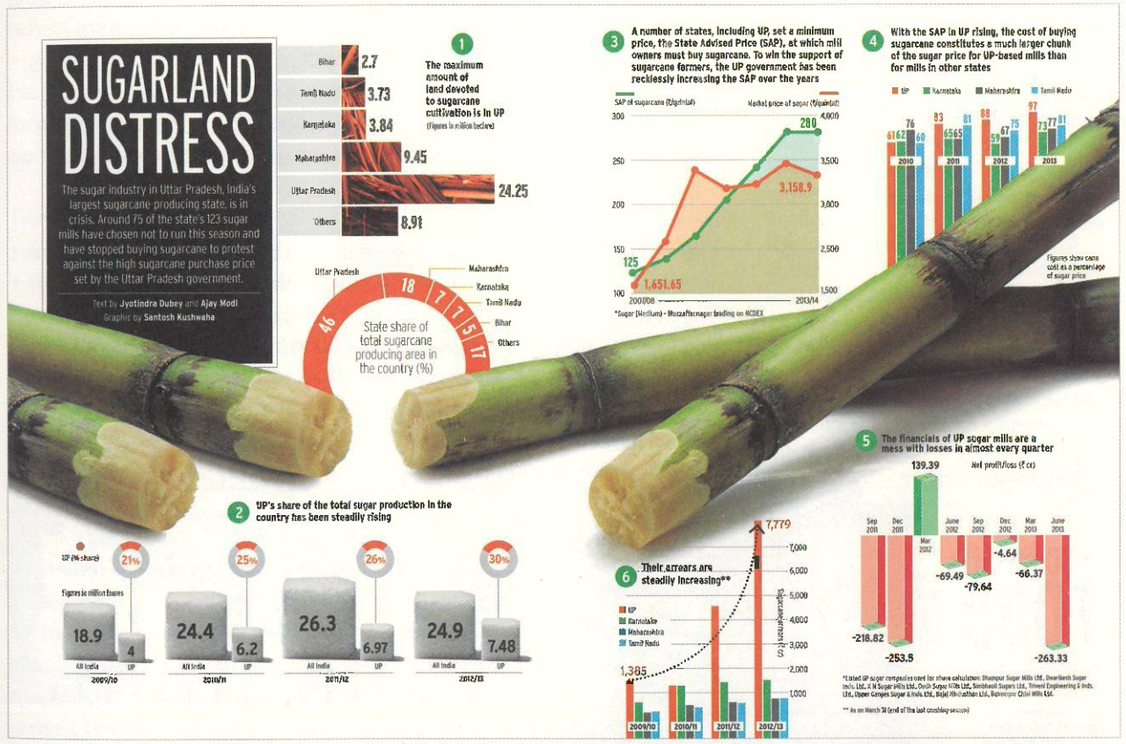
Similarly, in order to analyze China's rice production, the data will not be complete if it only uses data of major rice-producing regions in the South and ignores the statistical data of rice production in the Northeast, North China and other small areas.

If all data is not available—say, for example, because the government has not disclosed it—the designer will need to change the report strategy by narrowing the coverage. In the case of rice production in China, if it is difficult to obtain the rice yield data of the northern provinces, the designer can change the subject to 'Statistics on the development of China's major rice-producing areas.'

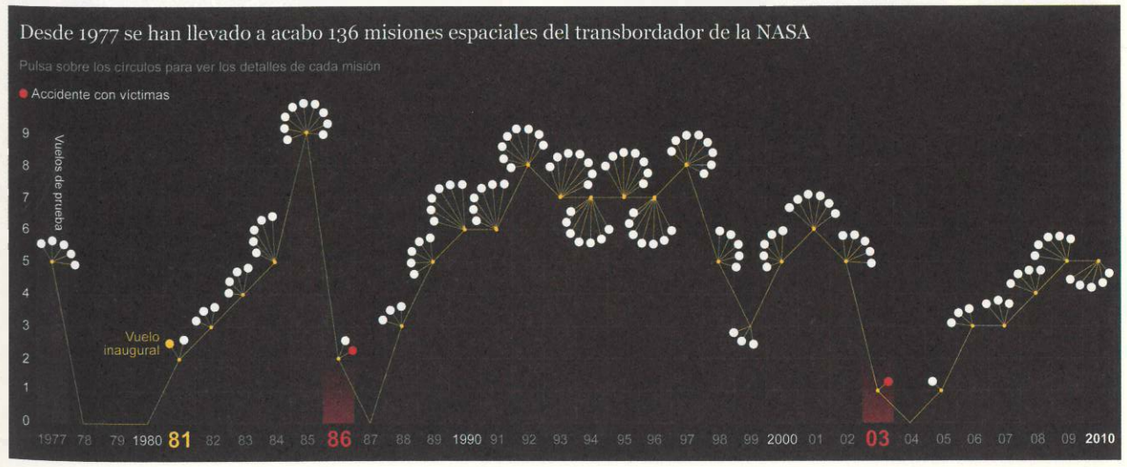
1. Line Graphs
In the above example, a line graph is used to describe the price trend of sugarcane in India from 2007 to 2013, with a green line representing the government-controlled prices and an orange line representing the prevailing market prices. The graph clearly shows the price fluctuation over the seven years (highest price, lowest price).

In addition to describing different data for different years, a line graph can illustrate trends over time. Figure 004 shows NASA's aircraft missions from 1977 to 2010. Due to the 33-year time span, it is appropriate to use a line graph to show specific flight missions and trends relating to the annual amount of missions over time.

003



004



005

Discomfort Food

Project Description /

The story is about increasing food inflation over several months. The reasons for the persistent high food inflation of the last few years have been much debated. There is a great need to improve supply chains and reduce wastage. Designer Santosh Kushwaha's popular graphic style uses real images. He thought it would get more traffic if he created something eye-catching using real images, rather than vector graphics. He used major food elements to create a human story. The story is serious, so he tried to express that in his visual representation.

Client / *Business Today*, *India Today* Group
Design Agency / Visual Best
Designer / Santosh Kushwaha
Completion / 2012

Line graphs are suitable for periodic data, such as an analysis of social media users, in which the graph can show the age distribution trends of social media users.

The following is a list of data types best illustrated by line graphs:

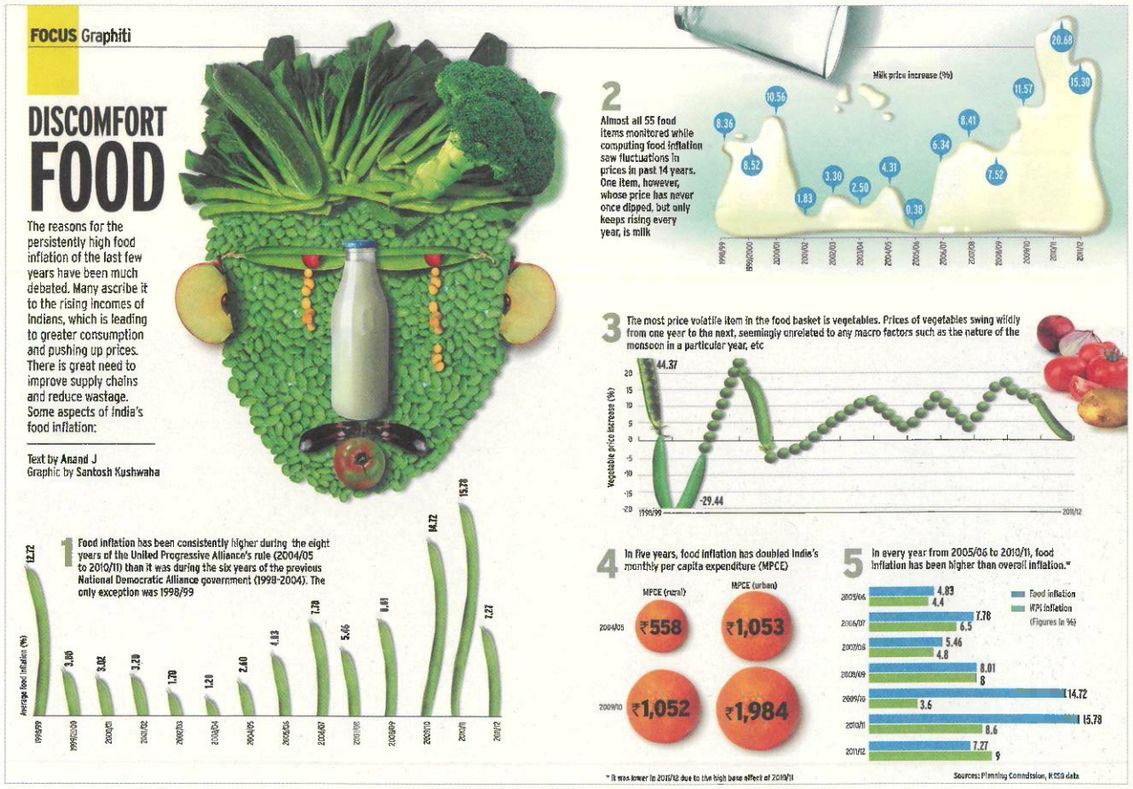
- Data with time features—for example, weather trends over long periods, monthly oil price fluctuations, and quarterly per capita consumption.
- Data that ranks numbers—for example, the age distribution of radio listeners or the salary distribution of various jobs.
- Data showing regularity—for example, temperature distribution from south to north, or the distribution of soil fertility according to altitude change.

Line graphs have the advantage of clearly illustrating trends. Figure 005, made up of bar charts and line graphs, shows the rising trend of food prices. Large fluctuation trends are best presented in bar charts or line graphs; small fluctuations are best illustrated using line graphs.

How to make a line graph:

- Determine the horizontal and vertical coordinates. Generally, variables such as data, age, and altitude are used as the horizontal coordinate. The vertical coordinate is the value of the data to be described, and the range of those values is determined by the maximum and minimum values of the data.

005



- Mark the data. Determine the position of the data in the coordinates according to the data values on the horizontal and vertical coordinates. In line graphs, one data value corresponds to one coordinate point.
- Connect the coordinate points. The amount of coordinate points directly reflects the size of the data sample.

2. Pie Charts

In the case of Figure 003, the pie chart is used to reflect the sugarcane yield of different areas in India, and Utta Pradesh is the largest sugarcane-producing area, accounting for 46% and occupying nearly half of the diagram. To make it more attractive, the pie is designed as a sector diagram, rather than a circle. Both pie charts and sector diagrams use area to represent the proportion of different data types and can directly reflect information, such as areas, regions, and percentages.

Pie charts are most useful for complete data sets. When the data set is incomplete, use the designation 'other' to label the segment that makes up the percentage shortfall.

Two features of pie charts:

- Integrity
- Parallel relationship between subset data

Figure 006 describes the four main causes of deforestation: agriculture, timber (the logging industry), pasture (cattle ranches), and fuel. These relate to the four main uses for trees—agriculture, industry, animal husbandry, and energy—which have a parallel relationship with each other and no overlapping relationship. When combined, they constitute a complete set.

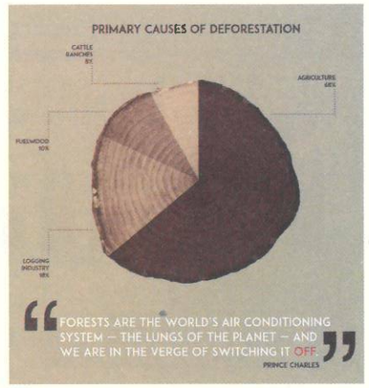
But when relating to other data, such as the Purchase Management Index (PMI), pie charts or sector diagrams may not be suitable because of the sheer number of indexes the infographic needs to cover. The PMI is known as the 'health checklist' for measuring a country's manufacturing industry, and has a number of indexes: production, new orders, commodity prices, inventory, employees, order delivery, new export orders, and import. Though these indexes cannot all be listed, the designation 'others' can be used to label the segment that makes up the percentage shortfall.

Pie charts can be made into flat round cakes, solid round cakes, single-layer circular cakes, multi-layer circular cakes, and sector diagrams. Whether a circular or sector type is used, the total area of the graph represents a complete set of data (which must total 100%), with each segment representing a percentage of the whole data set.

3. Bar Charts and Histograms

Bar charts and histograms look similar but they are not the same thing.

Bar charts use the height or length of bars or columns to represent data values. They generally contain data relating to different entities, with at



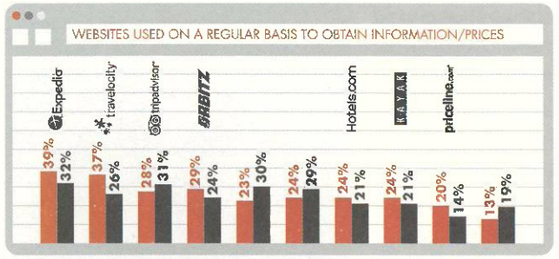
006

Deforestation

Project Description /

Unfortunately, deforestation is a rapidly increasing problem that negatively impacts the human race, as well as other forms of life in our environment. This poster summarizes the causes of deforestation, as well as the effects on the ecosystem. Also, it invites the viewer to take action to create awareness and to minimize the problem by doing simple things like taking care of trees and spreading the message of environmental protection. One of the main design elements of the poster is the tree trunk used as a pie chart. This imagery directly relates to the issue; it has been used to make a stronger impression on the reader, to give the piece an emotional impact rather than just showing numbers and graphics.

Designer / Regina Torres
Completion / 2014



007

007
The Portrait of
American Travelers Study

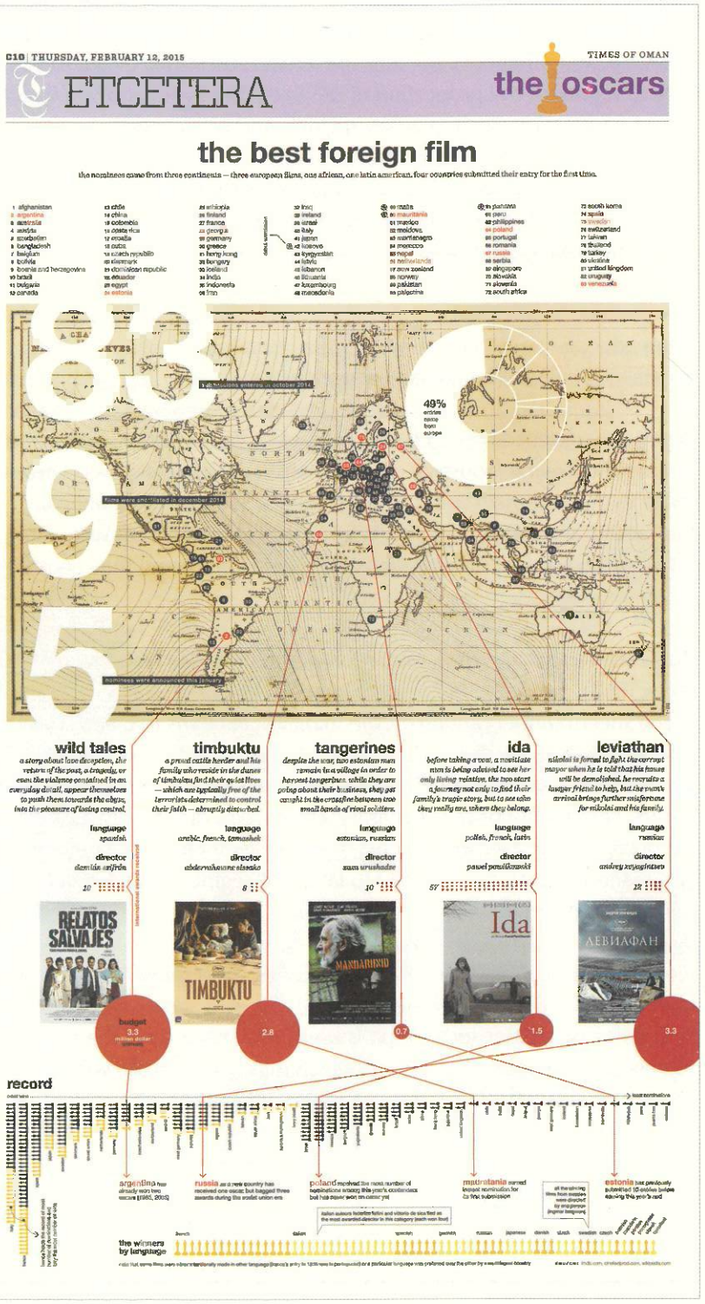
009

2015 Oscar Series: the Best Foreign Films

Project Description /

The infographic charts how many entries there were at the Academy Awards in 2015—from 83 countries to the nine short-listed films to the eventual five nominees. It also shows how many awards these films have won at other film competitions.

Client / *Times of Oman*
Designer / Adonis Durado
Completion / 2015



009

least two types of data for comparison. A gap is used between bars to indicate that they are separate entities, and different types of data should be different colors.

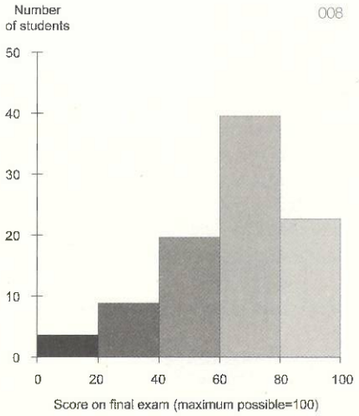
In Figure 003, five types of bar charts are used. The first chart describes the proportion of land growing sugarcane in the local agricultural area, with the length of bars representing the data value so that comparisons can be made between different sets of data. In the case of fewer groups and large data values, bars and columns can be distributed horizontally rather than vertically.

The data used in bar charts does not need to be a complete set because they are designed to compare different groups or entities. For example, Figure 007 shows the proportion of online visitors from the USA who search for travel information and prices, and the websites they visit. However, it only lists 10 popular websites, those that account for the highest numbers of people searching for travel information and prices rather than listing all the websites. This is the advantage of bar charts over pie charts and line graphs, which require complete sets of data.

Histograms differ from bar charts by using area rather than height to indicate quantity. The variables of the histogram must be continuous and can be divided into a number of equal sections, with each section's value determined by the following formula:
(Maximum value − minimum value) / group number = section value

In a histogram, there is no space between adjacent blocks, and the total area of the blocks—their height multiplied by their width—must add up to 100% of the total data set. Histograms not only enable the reader to conveniently observe the proportions of each element, they indicate the proportion of the whole that each block represents.

Figure 008, for example, shows the distribution of student numbers according to their scores. The horizontal axis has been divided into five equal parts, each representing a score range of 20 marks, with the size of each block also representing the proportion of students of the whole who received marks within that range.



008

4. Scatter Plot

Scatter plots can show the distribution and properties of all cases. Figure 009 lists all 83 films entered into the Academy Award competition for Best Foreign Language Film, with nine films screened as candidates and five chosen as the final nominations. All 83 countries are marked on the map, and the five nominated films are described in more detail. The biggest feature of scatter plots is giving prominence to individual cases.

Scatter charts present the data in the form of a set of points whose values are represented by their positions on the chart. The position is controlled by two variables. Scatter plots are often applicable to cross-category aggregated data—that is, the variables of horizontal and longitudinal coordinates belong to different categories. A benefit of

010

The London Olympic Games—Game Over

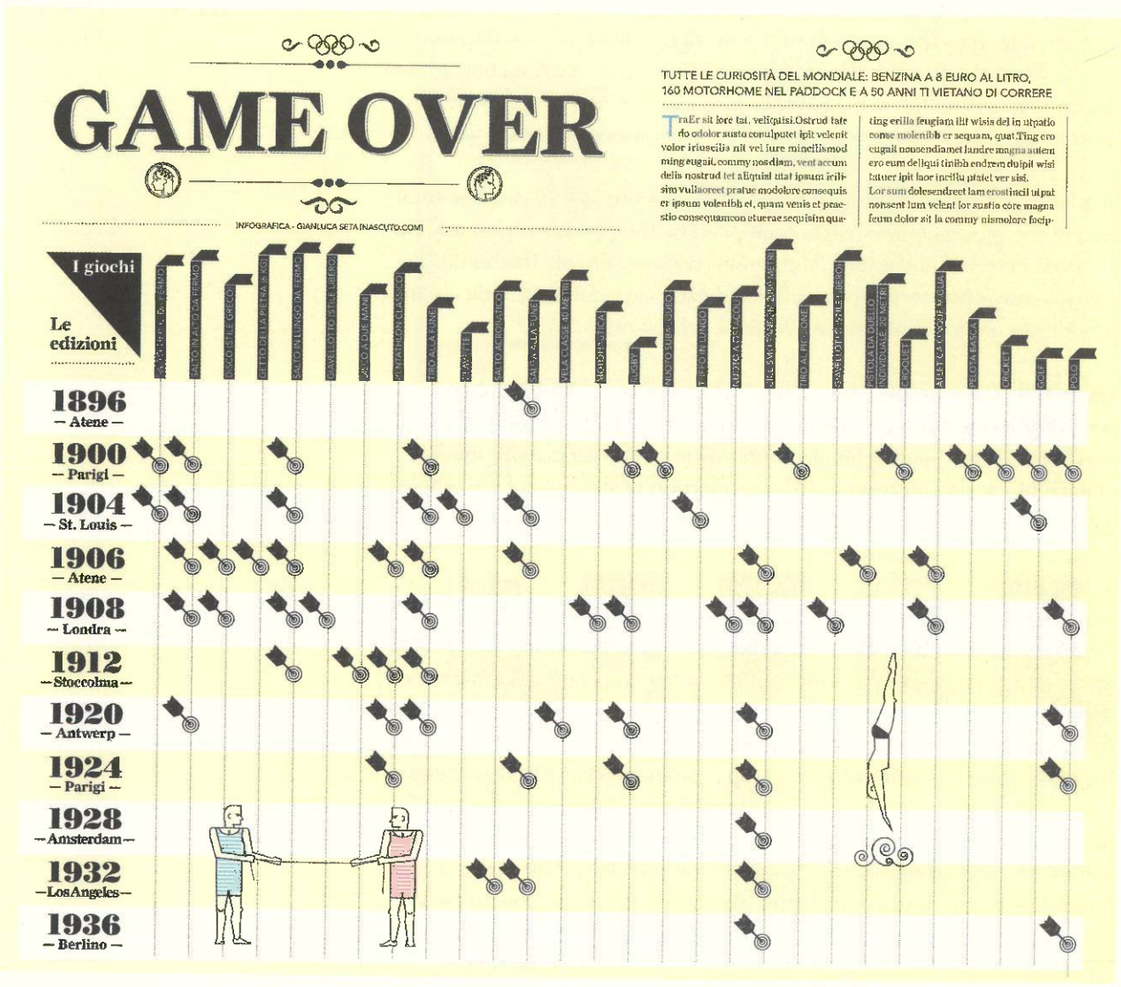
Project Description /
This overview features the strangest and most ancient games of the Olympics, which are no longer included in the modern games. The designer made use of a scatter plot to clearly explain the information.

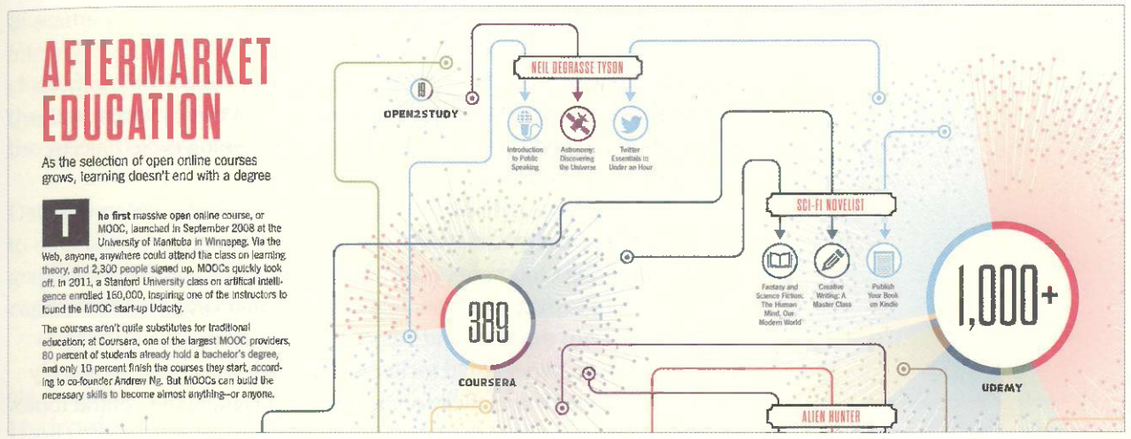
Client / *Sportweek—La Gazzetta Dello Sport*
Designer / Gianluca Seta
Completion / 2012

scatter plots is that they can show simple correlation, making it easy to see the scattered distribution or aggregation of the variables in two-dimensional space.

Making a scatter plot is similar to making a line graph. The difference is that data points of line graphs may not be detailed cases and the longitudinal coordinates represent the data values of individual cases, while scatter plot is still one of the variables of the data. There are only two steps for making a scatter plot: First, determine the two variables of the case, namely the abscissa and ordinate (there should be at least two variables like the latitude and longitude on a map); then, determine the positions of each case according to their coordinates. In Figure 010, the coordinates of each target are determined by two variables, which are 'ancient sport' and 'year'. The scatter plot reveals the gradual disappearance of the most ancient and competitive Olympic sports; from 1936 only two ancient sports were retained.

010

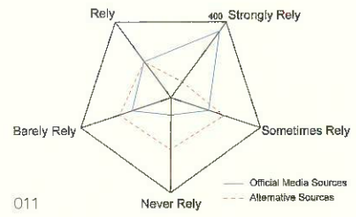




012

5. Radar Chart

A radar chart is also known as a web star, star chart, spider chart or irregular polygon. It is mainly used to compare data of relatively independent categories. Several equiangular spokes start from the same point, with each spoke representing an independent variable. Each spoke has a data point, from which the length to the center point is determined according to the proportion of the variable to the total amount. Connect these data points and a radar chart can be drawn.



011

For comparing the progress and qualities of each subproject, a radar chart is very direct and effective, using positions to describe distribution trends. Figure 011, for example, represents the statistics of news sources on which Chinese social media users rely. The blue star protrudes towards the northeast, meaning users strongly rely on official media sources. But the red star for 'alternative sources' does not show a clear tendency, as the star only shows a concave in the northeast but it's uniform in other directions.

There are two key steps for making a radar / star chart:

- Determine the number of spokes according to the requirements of data indicators. For example, in Figure 012 in 'Online Courses', the numbers and categories of free online courses are presented in a graph combining pie charts and star charts, in which pie charts show a rough category of all online courses, dots in pie charts represent the courses in each category, and the number of spokes represent the number of courses.
- Determine the positions of data points. Each spoke of radar / star charts represents sampled data and has a data point on the spoke whose length is in proportion to the size of data value. In the 'Online Courses', the lengths of spokes represent the course's popularity—the longer the course, the more popular it is.

Radar / star charts can be made with the aid of Origin, Visio, AI, and other software.

012
Aftermarket Education: the MOOC Landscape

Other Descriptive Methods

Descriptive data can use a variety of graphics, individually or jointly, to help interpret the data. With rapid development of the technology, the presentation methods of descriptive data are becoming more diverse, interesting, vivid, and lively.

1. Word Cloud / Tag Cloud

Word Cloud / Tag Cloud is a type of visual graph that uses Internet technologies to extract and randomly arrange some keywords from the text. It determines the fonts and colors of these keywords according to their frequencies. Word cloud technology can aid the reader to quickly grasp the core words and their frequencies, highlighting the central topics and related framework of the text.

Figure 013 is a BBC word cloud analysis of Indian leader Narendra Modi's Independence Day speech in August 2014. The graph shows that the most frequent words are 'India', 'one' and 'country', which are marked in brown. Other key words include 'government' and 'want'. Based on this analysis, the BBC discerned that India's leader was very eager to announce his ideas for the government and future of India.

The following websites provide word cloud automatic generation technology services:

- www.jasondavies.com
- www.wordle.net
- http://worditout.com

Word cloud technology is mostly used for analysis of text with a lot of opinion and a large number of words, especially when the central idea of the text is not immediately clear. Word cloud analysis can help confirm whether specific words appear frequently. It is often used for analyzing political leaders' speeches, refining the themes of websites, and illustrating government reports.

2. Mixed Graphs

In order to visualize some data news, it's sometimes necessary to use mixed graphs, which are suitable for presenting complex data that have a variety of indicators, variables and categories. Figure 014 shows Brazil's irrigation conditions from various standpoints through a combination

013

of scatter plots, line graphs, pie charts, maps, and other charts. In this infographic, the line graphs show a downward trend year by year, the pie charts highlight the proportion occupied by each economic sector, and the scatter plots clearly show the significant effect of water shortage on the whole country of Brazil.

Data journalism uses data as the main information carrier, giving priority to data presentation rather than text content. Sometimes one type of graph cannot cover all aspects of a story, so mixed graphs are used to capture all components.

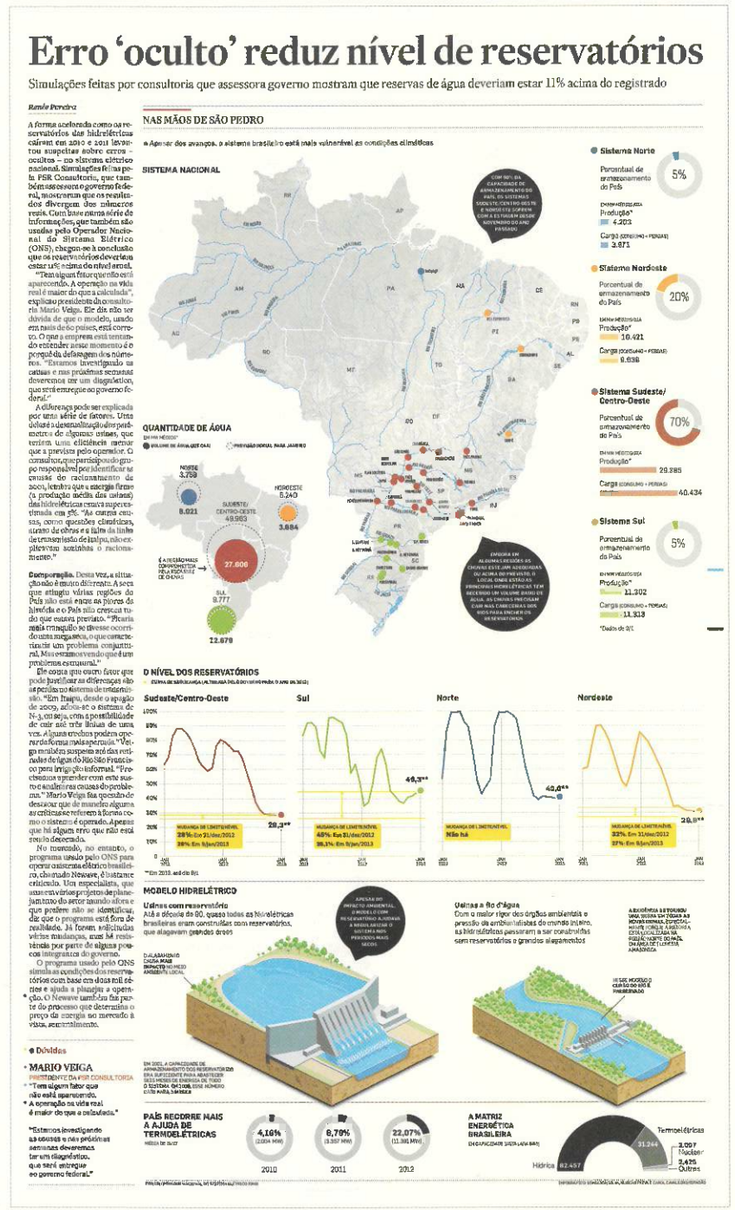
014

Hydroelectrics in Brazil

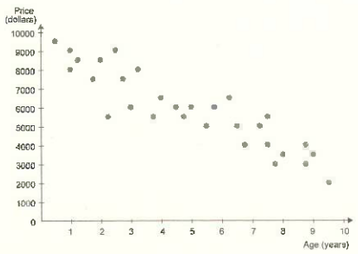
Project Description /

Hydroelectrics are Brazil's biggest source of power. However, climate changes make it rain less than usual in some regions, causing blackouts in some states. This infographic shows how Brazil's integrated power source works. The lack of rain in one region affects the whole country. Despite the southeast region producing more than 70% of the hydroelectricity, it's still deficient.

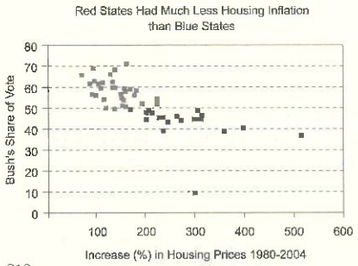
Client / *O Estado de São Paulo*
Designers / Carol Cavaleiro, Rubens Paiva, Edmilson Silva
Completion / 2013



014



015



016

This kind of work can be carried out with Microsoft Excel. Several types of graphs can be combined to make the data readable, visible, and clearly identifiable. Creating a data journalism graph is different from making a statistical chart, because the latter is used for scientific research and needs to explain the data in an accurate, rigorous way by using neutral, objective words. For example, '48%' cannot be interpreted as 'nearly half'. Data journalism charts, on the other hand, aim to report accurate data, but also contain some subjective opinions of the journalists. This type of chart promotes mass communication, enables more people to understand the hidden meanings in the data, and increases the practical significance of the data. Related pictures are often inserted into the graphs to make them more visual and direct.

Multidimensional Data Journalism Design

Correlation is the uncertain quantitative relationship between variables. With correlation data, if one variable changes, the other variable will change accordingly. News has an important function as a social watchtower, which not only describes what happens in brief words, but also explores the hidden factors, causes and effects of the event. It can also make predictions about the future. Data journalism can effectively work as a social watchtower, because the fundamental objective of correlation analysis and multidimensional multilevel analysis of the data is explaining and predicting the relationships and trends between variables.

1. Simple Correlation

If the data has a simple linear correlation, one variable increases/decreases in line with the other variable. Drawing a scatter plot can simply verify the correlation of two variables.

Scatter plots can not only present single cases, but also show the correlation between two variables, and if a correlation exists, the point distribution will be aggregated or scattered according to certain rules. Linear aggregation is linear correlation. For example, Figure 015 shows the points distribution of data in linear correlation; it surveyed 30 vehicles that were in service over different time periods, and marked a point for each vehicle according to their service lives and market prices. The graph shows that the market prices of vehicles decline along with the increase in service lives, which is a negative correlation.

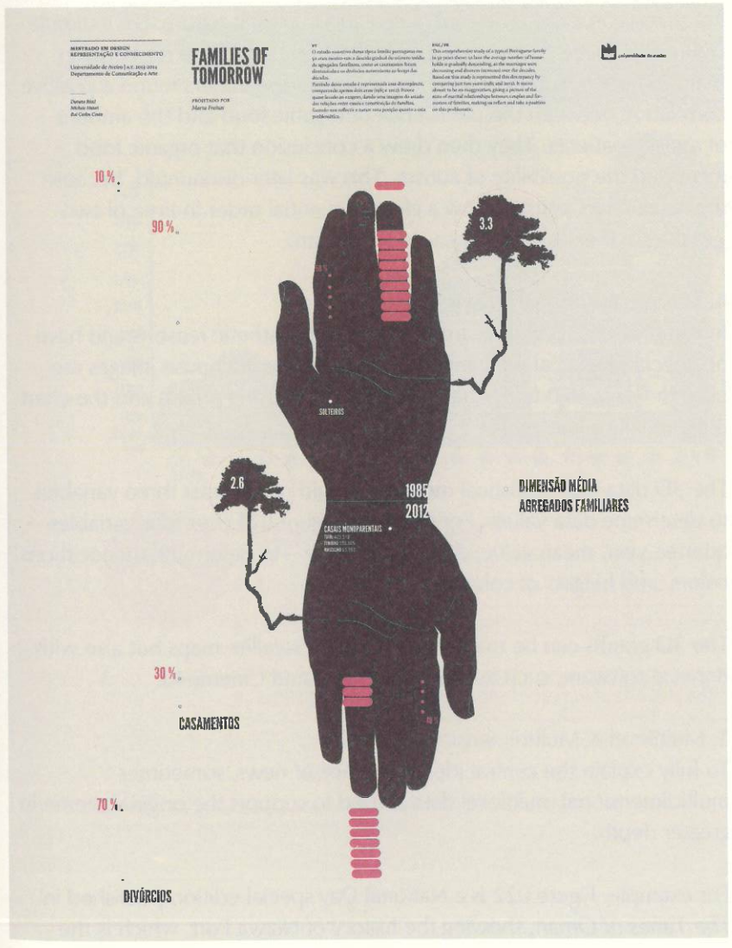
If scattered points are marked in different colors, individual cases can be compared. For example, Figure 016 represents the presidential electoral result of the former US President George Bush, with red color points representing the Republican vote and blue representing votes for the Democratic Party. The chart shows that Republican voters are mostly located in areas with lower house price inflation. Therefore, it can be said that the inflation rate of house prices is related to the votes for Bush.

2. Side-by-side Comparison

Data news often compares two data sets to highlight their differences, which can be used to create charts with strong visual impact. This kind of data should be presented with the following principles in mind:

- The same indicators should be used for comparison. For example, you cannot compare the variable of age in one data set with the variable of gender in another data set.
- The data sets should be comparable. Otherwise, the comparison will be meaningless.
- The difference between data sets should have social significance. For example, although the populations of China and the United States vary widely, comparing their consumption habits or per capita income levels has more social significance than simply comparing the two country's populations. That is to say, journalists should select some meaningful indicators for data comparison.

Here is an example. Figure 017 shows a comparison of the numbers of married and divorced households in Portugal in 1985 and in 2012. The graphic reflects a sharp rise in the divorce rate as well as a fall in the marriage rate, reflecting a main problem of the current society in Portugal. This chart is concise and to the point, with obvious comparisons and profound social significance.



017

017

Families of Tomorrow

Project Description /

This comprehensive study of typical Portuguese families over 50 years shows how the average number in a household is gradually decreasing, as marriages decrease and divorces increase over the decades. Based on this study, two years in particular are compared (1985 and 2012). The results are startling. They almost seem like an exaggeration, causing us to reflect on the problems surrounding marriage and the formation of families.

Designer / Marta Sofia Ribeiro Freitas
Completion / 2013

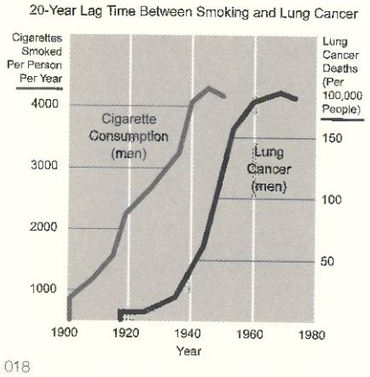
020

Our Changing House Market

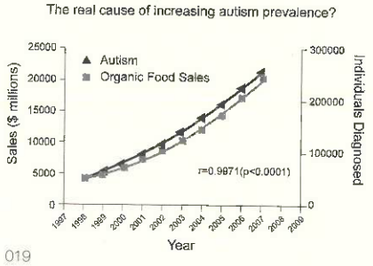
Project Description /

The client wanted to show data of 40 years of sale prices and stepped growth. Such data can be very dry, so Rahadyo Widyastomo needed to present the data visually in an interesting way to engage customers. The data is blended with the graphic of a house.

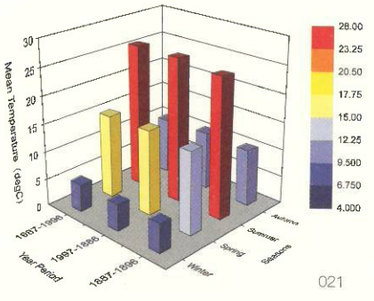
Client / mitchellhomes.com
Designer / Rahadyo Widyastomo
Completion / 2013



018



019



021

3. Causation

As news needs to explain social phenomena, data is more convincing than expert opinions in the exploration of causation, because data carries objective, unchanging facts, while expert opinions are often subjective and open to external influences. As a result, audiences and journalists tend to rely on data to present neutral and objective explanations.

The data with following conditions can be assumed to have a causal relationship:

- Variables of the data can be divided into two or more sets. One of these sets is the independent variable, which is assumed to be the causal variable; and the other set is the dependent variable.
- Dependent and independent variables have a sequential order in time.
- Independent variables cause changes in the dependent variables.

For example, Figure 018 shows a famous study arguing that tobacco causes lung cancer. From the chart we can observe that tobacco consumption is an independent variable in terms of time, and it was the sharp rise of tobacco consumption in 1900 that resulted in the increase of lung cancer deaths 20 years later.

However, correlations are not necessarily causal in nature. For example, Figure 019 shows a case of the identification of a causal relationship that did not, in fact, exist. By chance, some researchers found a positive correlation between the prevalence of organic food and the amount of autistic patients. They then drew a conclusion that organic food increased the possibility of autism. This was later discounted, because the researchers neither show a clear sequential order in time of two variables nor exclude other causes of autism.

4. Three-dimensional Graph

In some cases, 3D graphs are used only for aesthetic reasons and have no special statistical meaning. For example, the 3D house images are used in Figure 020 to illustrate the subject (housing prices) and the chart is essentially a line graph.

The 3D data with statistical meaning should use at least three variables to determine data values. For example, Figure 021 uses four variables—quarter, year, mean value, and temperature—to determine the locations, colors, and heights of columns.

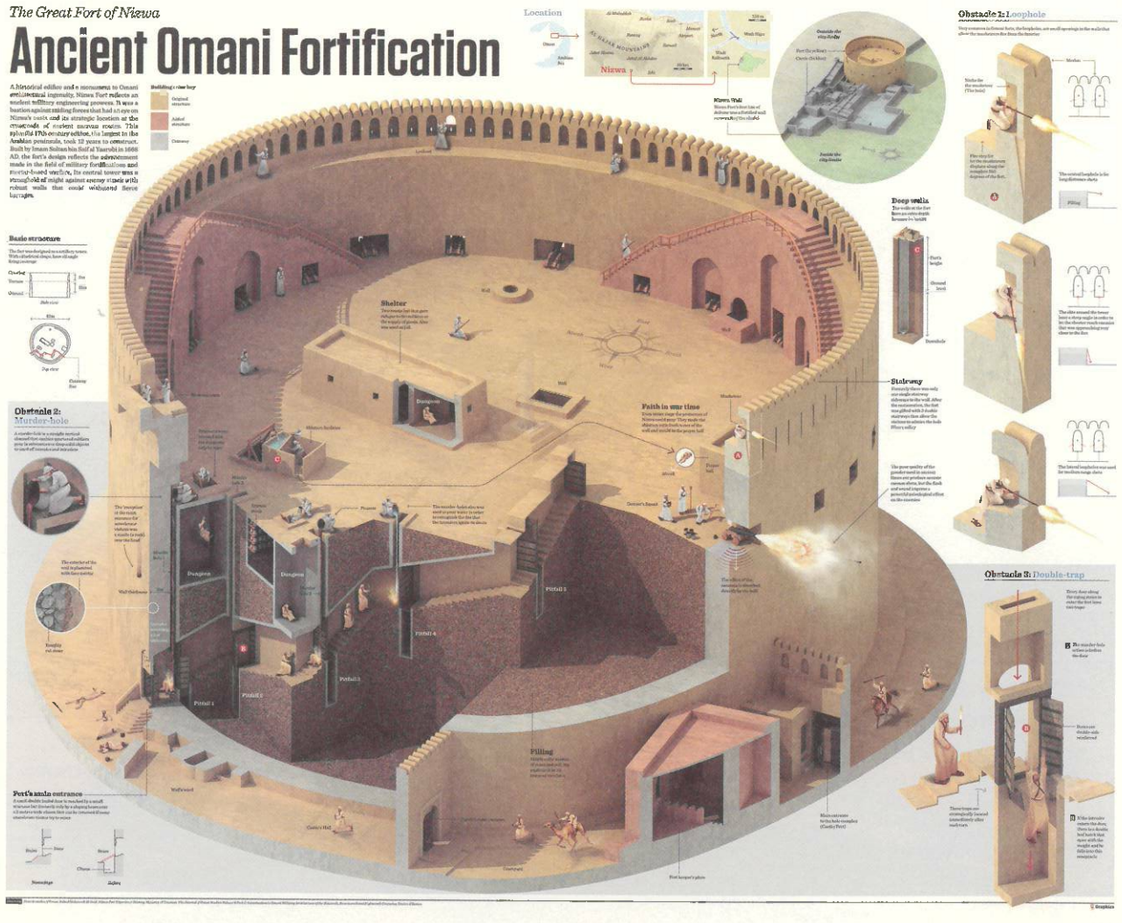
The 3D graphs can be made not only using satellite maps but also with drawing software, such as 3ds Max, Maya, and Cineme4d.

5. Multilevel & Multidimensional Data

To fully explain the central idea of a piece of news, sometimes multidimensional multilevel data is used to support the original theme in greater depth.

For example, Figure 022 is a National Day special edition published in *The Times of Oman*, showing the history of Nizwa Fort, which is the

020



022

The Great Fort of Nizwa

Project Description /

Antonio Farach was commissioned to produce an infographic about the forts in Oman, to be published in the newspapers *Times of Oman* and *Al Shabiba* as a double page for Oman's National Day celebrations. He decided to focus on Nizwa Fort because it is one of the most majestic among the more than 500 forts that still exist in Oman, and because of its unique zig-zag trap-stairs system.

Client / *Times of Oman*
Designer / Antonio Farach

oldest and most representative castle in Oman. The 3D sectional view shows the reader the inner structure of the castle, including the unique Z-shaped stairs. The functionality of each part of the castle is vividly described in text and strong images, giving the final work great collection value.

Interactive Data News Design

Interactive news uses software to analyze and interpret large volumes of constantly changing data. Common interactive graphs include real-time maps, process flowcharts, and heat maps that show intensity.

1. Real-time Maps

The most common type of interactive data news is the weather forecast, which achieves real-time broadcasting with the aid of satellite technology, showing the flowing states of clouds and cyclones. In addition to weather and geographical conditions, major international events and news stories involving a large amount of statistic data may be presented in the form of real-time maps. For war reports, for example, TV news often show the satellite maps of war zones, accompanied by descriptive data.

From a technical point of view, it's very difficult to create a real-time map that needs to monitor a vast amount of constantly changing data and requires the assistance of data mining technology. This creates fresh challenges for journalists, because the work on deep data mining and big data analysis requires computer science knowledge.

2. Process Flowcharts

Real-time dynamic graphs are often used to illustrate the ups and downs of the stock market. They show data changes and developments in a real-time manner, similar to live TV broadcasting.

The UK-based *Financial Times* often launches interactive data news in its Interactive news column, in which it often features graphs about stocks and bonds. These graphs will show additional data analysis and relevant news when the readers move their cursor to certain time points.

The process flowchart is static in the print media. Figure 023, for example, uses text, numbers, patterns, charts, and other elements to present the migration of the monarch butterfly from Canada to Mexico as well as its evolutionary history and hatching process.

These graphs help journalists explain a subject's dynamic features over a specific time period, so they can present a good overview and further explain their findings.

3. Heat Maps

Heat maps highlight data density or data size with color brightness or geometric symbols, and are often used for news about diseases, disasters, and temperature. Dynamic heat maps can monitor the distributions and trends of information in a real-time manner.

023

Monarch Butterfly from Canada to Mexico

Project Description /

This infographic forms part of a special project '40 years of the monarch butterfly in Estado de México'. It focuses on the monarch butterfly's travels from Canada to Mexico, and the evolution from a caterpillar to a beautiful butterfly. The infographic also presents the characteristic of the two sexes, and the food and nutrients needed to support the long journey.

Design Agency / Diario La Razón de México
Designer / Diana Estefanía Rubio

024

Dilma Won

Project Description /

The results of the presidential elections in Brazil won by Dilma Roussef are illustrated here in a simple day-by-day format. This type of work usually does not reflect the reality of the vote in each part of the country, but the designer believes this map is the best way to effectively interpret and display the election results, as it shows who had the most votes in each municipality according to its size.

Client / *Diário Económico*
Designer / Mário Malhão
Completion / 2014

Heat maps are often used to present meteorological data, such as temperature, air pressure, and humidity. Graphs that combine heat maps and real-time maps have a larger visual impact. Taking Figure 024 as an example, the votes for Brazil's presidential election are shown in the heat map, which does not directly present the actual numbers of votes but effectively predicts the voting results of each state. As the heat map predicts, Dilma Roussef won the final election.

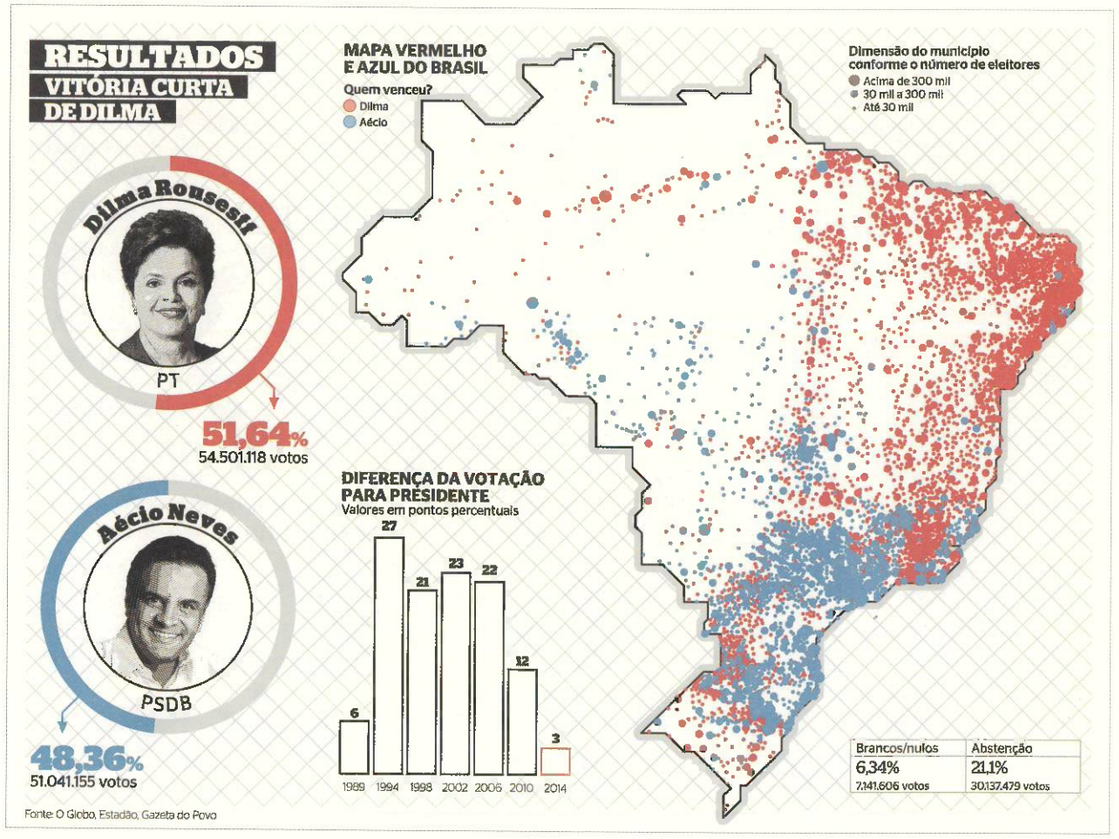
Data is the most precious resource of interactive news. A piece of interactive data is a small database, which is open to the reader in the graphs drawn. The reader can get the data by simply moving his/her mouse and keyboard. For example, Figure 025 is a piece of data news launched by *The Los Angeles Times*, whose data source is teachers' assessment data from 11,970 elementary schools in Los Angeles. From here, the readers can search for assessments of their primary school's teachers.

The simplicity of this kind of data presentation method is that readers can easily obtain the data they're after, and journalists do not have to analyze and classify the vast amount of data piece by piece.

4. Others

With the help of the Internet, dynamic data is presented in a variety of ways in social news and science news. Common data types suitable for

024



dynamic graphs are crime rates, traffic conditions, disaster casualties, and regional distribution of disease outbreaks.

Dynamic data is usually presented in multilevel and multidimensional dynamic graphs, based on descriptive diagrams and using a variety of presentation methods. One of the common methods is maps. For example, Figure 026 is the dynamic graph of the bicycle accident distribution in San Francisco, from which the readers can learn about the total bicycle accident numbers in some hot spots by clicking these spots. This data graph is a combination of multiple presentation methods such as a map, a scatter diagram, and a heat map.

When making data charts, news agencies and journalists inevitably set some frameworks and impose some subjective intentions to the objective data. For example, Figures 027 and 028 show the percentage of American families subscribing to daily and Sunday newspapers in the past 12 years, and the inevitable drop due to the emergence of the Internet. In Figure 027, the drop in subscriptions does not look significant. But when the vertical scale is narrowed to 30%–60% in Figure 028, the changes are made to appear much more dramatic. A journalist who wants to emphasize the changing media landscape might choose to use Figure 028.

The work of data journalists is scientific work to some extent, and the processing of the data has to follow some basic principles:

(1) Truthfulness

Journalists should ensure that data collection has been rigorous and authentic, and data sources should be formal and reliable.

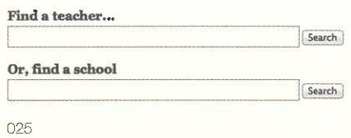
(2) Integrity

In the process of presentation, description, and explanation, the integrity of the data should be maintained without overgeneralization.

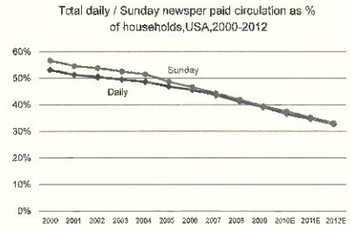
(3) Rationality

The explanation of data correlation and causation should be scientific and rational, without misusing data correlations.

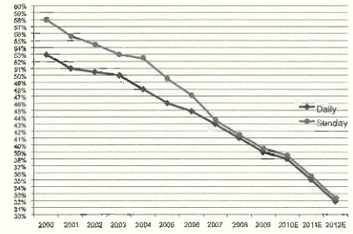
025



027



028



026