# COMP6234 Data Story

Thanadon Fuengworatham
*School of Electronics and Computer Science*
*University of Southampton*
Southampton,United Kingdom, SO17 1BJ
Student Id: 29392802, Email: tf2n17@soton.ac.uk

***Abstract*—Terrorist attacks have been a big issue of the world and getting riskier to our daily life. This data story is aiming to deliver insights of the data to the readers in interesting way with solid backup from underlying statistics. The main questions that will be raised up in this data story are "Where were the favourite targets of terrorists?", "Why they chose those places or cities?" and "What is the common criteria among terrorists to choose a target?". The answers of these questions would help us aware of the places which are likely to be targeted and, ultimately, be able to avoid those places to reduce the risk of being harmed by terrorism.**

## I. Introduction

The data story is the combination of data, visualisation and art. The report will start from the description of each dataset, followed by the theories behind the design of the story and the visualisations and will end up with the possible improvement of each element of the data story. To focus on current terrorism trend, this data story will focus on the period after the 9/11 attack on United State (09/11/2001), which is considered as the evolution of the terrorist. However, the visualisations enable interactive features to users where they can freely change the dates and specifically choose the areas of their interests.

## II. Data Sources

There are 2 datasets referenced in this story. First is world terrorism dataset which is used to presents the statistics of terrorism around the world including specific details of where the targets and what the characteristics of those targets. Second is population dataset. This data is used to support the assumption of whether city size is correlated with terrorist attacks. Follows are the details of each dataset.

### A. Global Terrorism Dataset

This is the history data of terrorist attacks which have happened around the world from 1970 to 2016. The data comes with 170,350 records and 135 columns with main information categorised in date, location, fatality and details of the attack, terrorists and their target. The date information is separated into 3 columns year, month and day. While location information is very specific, containing coordinate (latitude – longitude), province, city, country and region. The dataset is collected, managed and organised by University of Maryland and is annually updated. There is the document provided explaining the detail of each column, method of data collection and the motivation behind this project. Overall, the data come with complete information with standard data content in city name, country and date.

| Column | Data type | Sample |
|---|---|---|
| Day | Quantitative , Discrete | 11 |
| Month | Quantitative , Discrete | 9 |
| Year | Quantitative , Discrete | 2001 |
| Region | Qualitative, Nominal | North America |
| Country | Qualitative, Nominal | United State |
| City | Qualitative, Nominal | New York |
| Target Type | Qualitative, Nominal | Business |

Fig. 1. Data type and sample of Terrorism data [1]

*1) Data type and sample:* The raw data contains many fields; however, only essential columns, shown in Figure 1, are selected to create the visualisations.

*2) Data processing:* The dataset comes with good quality and structured format without need of data cleansing process. Only one problem is that the date comes in different columns. The data transformation is required on the dates data. Thus, Day, Month and Year columns are transformed to Date column, which is the combination of those three fields. The new column type is defined as date and the sample is "11/9/2017".

### B. UN World Cities Population Dataset

This dataset contains information of cities population around the worlds, collected and managed by United Nation (UN). The file is in CSV format with 51,941 records and 10 columns. The main columns are country, city, population and collected year in data type of text, text, number and number respectively. The data cover wide ranges of all cities around the world; however, the referenced year when the population data was collected is not consistent. For example, some countries are updated recently but some are not.

| Column | Data type | Sample |
|---|---|---|
| Country | Qualitative, Nominal | Russia |
| City | Qualitative, Nominal | MOSKVA |
| Population | Quantitative, Discrete | 11918057 |

Fig. 2. Data type and sample of Population data [**?**]

*1) Data type and sample:* Only three columns are used in the visualisation which are shown in Figure 2.

*2) Data processing:* Data cleansing process is required on the City column as the city name does not match to the city name in Terrorism data. The United Nation allows some original words to use in the city name. For example, instead of Moscow, the dataset use MOSKVA, which is the traditional city name used in Russia. This cause unmatched data between column City in Terrorism and Population dataset. The solution is to manually update the city name in Population data in accordance with the city name in Terrorism data. Fortunately, the manual updates are applied on only the top-6 European countries that are used in the visualisation.

## III. Data Story and Visualisation

The presentation in this data story is separated into 3 parts. Firstly, the story will present overview of the terrorist attacks around the world to make the reader understand the terrorism trends nowadays. This part is presented by world map which is achieved by using location data from the Global Terrorism dataset. Trends of terrorist attacks across different regions over the years are included to compare the trends between regions. Secondly, the story will narrow down into city level focusing on top-6 often-attacked cities in Europe. This is to point out the assumption of whether popular cities in a country are likely to be terrorist target. Lastly, the final part will dig down type of the target. The story will reveal what types of target such as police, government, military or business area were often attacked by terrorists. The visualisations in the data story are developed from Tableau 10 and designed according to some important visualisation theories which will be deliberated in following sections.

### A. Data story

The data story is deliberately designed according to following theories.

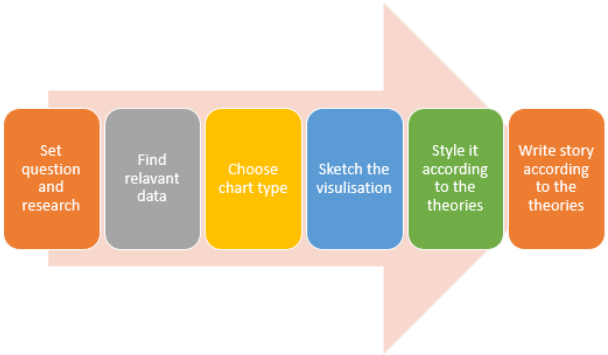


Fig. 3. Application of Cairo's methodology in development process [2]

*1) Cairo's methodology:* The concept of Cairo's methodology is applied in overall development process of this data story. According to Figure 3, the data story was started from one initial question of that "How can we avoid the harm from terrorism?" with some consequent specific questions such as "Where is often attacked?" and "Why?". Then many researches were acquired to correspond with the questions followed by ideas about the story then selecting relevant data to present in the story. After that the visualisations were sketched with the content outline to draft the story. This enable us to see the flow of the story and check if all of the questions are answered. Finally, last step is to write the story coordinating with the visualisation. Some theories in the lectures are referenced to control the quality of the story, especially, to manage the connection between the visualisations and the stories.

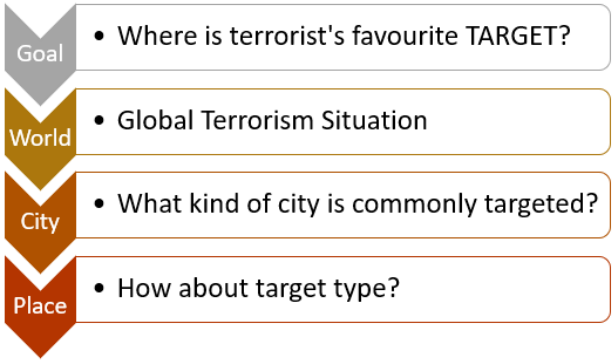


Fig. 4. Data Story design in Linear structure [3]

*2) Linear structure:* The story starts by throwing the ultimate question to the readers to shape their focus. Then the headlines in each section are planned to start from big target to smaller ones which are world, city and place respectively as shown in Figure 4.

*3) Martini and Drill-down structure:* [4] The data story balances the structure between Author-driven and Reader-driven with following theories: -

*a) Martini:* The story starts from author-driven and end with reader-driven with following plans.

• The story focuses on the 9/11 which is used to initialise the start date in the map to shape readers focus.

• The story picks up the statistics from all regions showing on the map then the map is initialised with all regions.

• At the end, the story ends with the question "How about the situation in your city?". This is designed to stimulate readers to use the interactive features in the map to explore the data of their interest.

*b) Drill-down structure:* The story presents the general idea of terrorism about world situation and evolution of terrorism from 9/11 attacks then enable full interactions in each visualisation to the readers. The drill-down features of each visualisation are designed as follows: -

• World's Terrorist Attacks dashboard - allows people to drill down the data from region to country level. This is designed to benefit individual preferences in case that the readers want to explore the data in specific region and country.

• Top-6 Most Attacked Countries in Europe dashboard - the bubble chart allows user to select individual city to see the most dangerous place specifically for that country.

### B. Visualisation

*1) Overall design:*

*a) Colour:* Red colour is used to represent number of attacks in each chart according to meaning of danger in the Colours in Cultures wheel in Figure 5. This is to create emotion to the readers.

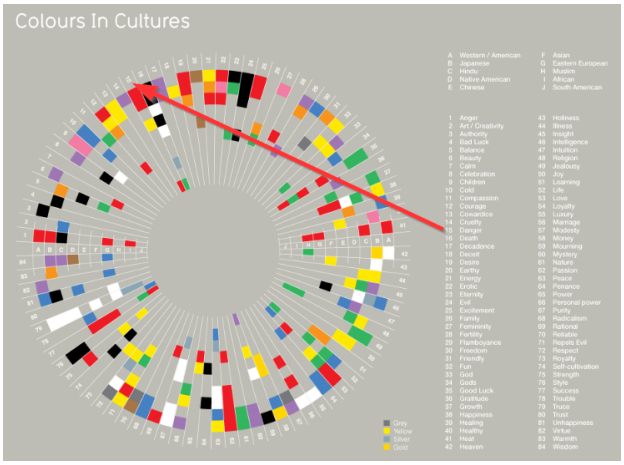


Fig. 5. Meaning of colour in different culture [5]

The theme of all visualisations is controlled to remain in the natural colours with red and blue as shown in Figure 6. This is to make the charts clearly visible to everyone including colour-blindness people.

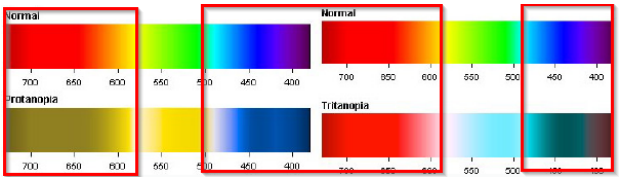


Fig. 6. Example of colour in colour-blindness perspective [6]

*b) Data to ink ratio:* According to Tufte's [7] theory, some charts in the story are designed to have more rooms for data by reducing the used of line and grid. For example, Most Targeted Countries bar chart and Trend of Terrorist Attacks line chart. However, the grid line on the bar chart presenting target type, in Figure 7, is not removed. This is because the chart height covers entire space on the left of the dashboard, which makes the top bar far from the X axis. Thus, it is better to have grid line to facilitate the estimation of the value from far distance.

*2) Top-6 Most Attacked Countries in Europe dashboard:*

*a) Population vs Number of Attacks (Bubble Chart):*

- *Key data*
  - Population data
  - Number of terrorist attacks
  - Filtered by country and city

- *Chosen reasons*
  Bubble chart is not good at comparing the data that need precision but it is used because the goal is to show the approximate comparison of population and number of attacks in each country. Thus, the main focus would be bigger or smaller for risk rate and dark or light blue for population, not specific quantity.

- *Avoided bias and confusion*
  Bubble chart is good at comparing multi-dimension data while it will be complicated if bar chart is used. For instance, relationship between population and number of attacks can be intuitively presented by colour and size of the bubbles, respectively. However, if we do the same in bar chart by colouring the bar with the meaning of population, it does not make sense to readers and also cause confusion.

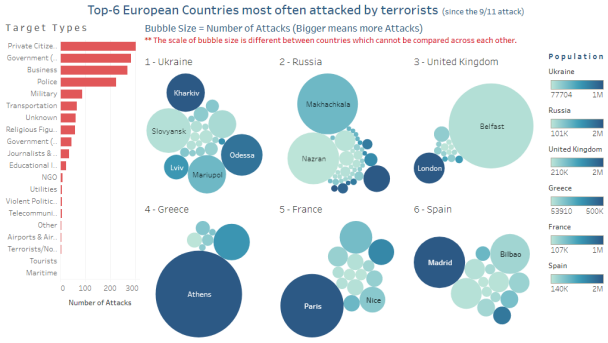


Fig. 7. Top-6 European Countries most often attacked by terrorists dashboard [8]

- *Answered Questions*
  - Is big or small city usually targeted?
  - What target type is often attacked?
  - What is the difference of target type in big and small city?
  - Where should be avoided when travelling in the risky cities?
  - What is the riskiest country in Europe?
  - What is the riskiest city in each of the countries?

*b) Target Types vs Number of Attacks (Bar chart):*

- *Key data*
  - Target type
  - Number of terrorist attacks
  - Filtered by country and city

| Rank | Aspect judged | Example |
|---|---|---|
| 1 | Position along a common scale | Scatter plot |
| 2 | Position on identical but nonaligned scales | Multi-panel graph |
| 3 | Length | Bar chart |
| 4 | Angle, Slope | Pie chart |
| 5 | Area | Bubble chart |
| 6 | Volume, Density, Color saturation | Heatmap |
| 7 | Color hue | Treemap |

Fig. 8. Ranking of perceptual task in Cleveland and McGill's theory [9]

- *Chosen reasons*
  - According to ranking of perceptual task in Cleveland and McGill's theory in Figure 8, bar chart is in rank 3, which is good at comparing data that need precision. As the gap between number of attacks in each city is small, it is better to utilise the advantage of bar chart to justify the differences. The reason that scatter plot and multi-panel graph in rank 1 and 2 are not chosen because the number of data point is small to be plotted in scatter plot, while the data have only 1 dimension which does not required muti-panel graph
  - According to the guideline in the Data Points book shown in Figure 9, bar chart should be used to answer question of "What is the best and worst?" which matches with the objective of this visualisation

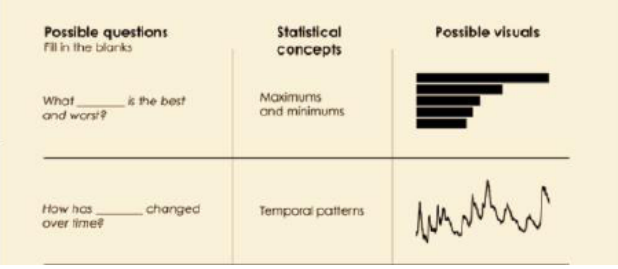


Fig. 9. Chart guidelines in Data Points book [10]

- *Avoided bias and confusion*
  At first design, the highlight table was used and number of attacks were presented by intensity of colour. This make it difficult for readers to determine small differences between target types. As a result, bar chart is instead used as it delivers numbers to readers in perceptual level, according to the Figure 8, which is more precise for this data. This avoid bias in small differences in numbers and makes more sense on this data.

- *Answered Questions*
  - What is the riskiest target type in each city/country?
  - What type of places should be avoided in the risky city/country?

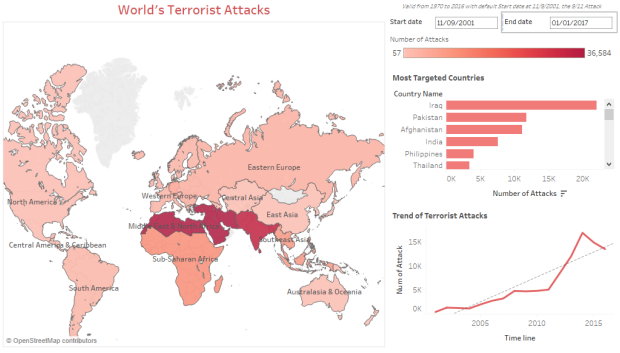


Fig. 10. World's Terrorist Attacks dashboard [8]

*3) World's Terrorist Attacks dashboard:*

*a) Interactive World Map (Filled Map):*

- *Key data*
  - Region
  - Number of terrorist attacks
  - Filtered by country and date

- *Chosen reasons*
  - The aim of this visualisation is to overview the overall picture of terrorism around the world and precision is not important. This make map best fit to the aim.
  - It is easier and intuitive for readers to explore their interesting area by map rather than other types of chart.
  - The risk in form of number of attacks can be presented by colour on the map. And the intensity of colour can be used to roughly compare risk between regions.

- *Avoided bias and confusion*
  As the ultimate goal of this visualisation is to allow user to interact with their region/country choices. To make it clear and easy to play with, map is chosen to present this data as it is the most understandable form for the interaction related to geographical data. This is intuitive to play with which can prevent readers confusion.

- *Answered Questions*
  Which area in the world are dangerous from terrorism?

*b) Most Targeted Countries (Bar Chart):*

- *Key data*
  - Country
  - Number of terrorist attacks
  - Filtered by region and date

- *Chosen reasons*
  The reason is the same as the Target Types Bar chart according to the Cleveland and McGill's theory and the guideline in the Data Points book

- *Avoided bias and confusion*
  Bar chart is used with the same reason as specified in Target Types.

- *Answered Questions*
  - What is the riskiest country of the world/in each region?
  - What is the top-6 most attacked countries of the world/in each region?. It is the design intention that fixes the height of the panel to display only 6 bars at the first place.

*c) Trend of Terrorist Attacks (Line Chart):*

- *Key data*
  - Year

- Number of terrorist attacks
- Filtered by date, region and country

- *Chosen reasons*
  The main objective of this visualisation is to show trends of terrorist attacks over the years. According to the guidelines in Figure 7, line chart is best fit with question "How has terrorist rate changed over the years?".

- *Avoided bias and confusion*
  - Line chart is chosen to deliver trends of data over times as it is more intuitive and clearer than other types of chart. For example, if bar chart is used, it can provide more precise number but it does not align with the main aim which is the ability to answer question whether the trend of terrorism increase or decrease over the year, not how much difference in number of attacks.
  - The dimension of date data is reduced from day to year level. As too many details could lead to confusion and the main aim is to just tell the trend. Thus, year is specific enough to achieve this task.

- *Answered Questions*
  - What is the terrorism situation around the world/ in each region/ country?
  - What is the terrorism trend? Increase or decrease?

## IV. Improvement

### A. Visualisation

- In bubble chart, the scale of bubbles size is specified by text. To improve this, it is better to show an example of bubble next to the chart as the scale measurement to tell different number in each size.
- In the map, highlight colour is shown in region level, not country level. This design is to show readers the overview of terrorism situation in the world. However, the disadvantage is that some safe countries that belong to the risky region will be presented in high risk area while it is low risk. To improve this, there should be an option for user to choose to show map in region, country or even city level in case the direct focus on city is preferred.

### B. Story

The story is more like describing the statistics from the past to the readers and warning them to avoid those risky places. However, the terrorism is evolving every year, it would be more interesting to show some predicted data. For example, allowing user to choose city and the visualisation will show the predicted result of probability that the terrorism will take place on each target type. This can answer more interesting question of "What will be the next target in this city?"

## V. Conclusion

Many theories and techniques are applied in this data story to deliver insight from the data to readers in the most effective way. This is achieved by balancing between author-driven and reader-driven style together with interactive visualisation and deliberately design of charts in terms of colour, size, data-to-ink and proper types. Taking all of these into account, not only the data story can effectively answer all of the important questions set-up at the headlines, but allows readers to dig down the interactive visualisation in the area of their interest which potentially leads to further discussion.

## References


[1] "Global terrorism database," (Accessed on 01/11/2018).
[2] A. Cairo, "The functional art."
[3] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations. in visual languages," September 1996, 1996. Proceedings., IEEE Symposium on (pp. 336-343). IEEE.
[4] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," 2010, visualization and Computer Graphics, IEEE Transactions on, 16(6), pp.1139-1148. Vancouver.
[5] "What colors mean in different cultures — visual.ly," (Accessed on 01/11/2018).
[6] "Colblindor — all about color blindness," (Accessed on 01/11/2018).
[7] Tufte, "The visual display of quantitative information book," September 1996, 1996. Proceedings., IEEE Symposium on (pp. 336-343). IEEE.
[8] T. Fuengworatham, "World terrorism data story," (Accessed on 01/11/2018).
[9] Tufte, "Cleveland and mcgill paper in journal of the american statistical association," 1984.
[10] N. Yau, "Data points: Visualization that means something," 1984.