

Innovation Action (IA)

ICT-14-2016-2017

H2020-ICT-2017-1

Enabling procurement data value chains for economic development, demand management, competitive markets and vendor intelligence



Deliverable 4.4

Visualisation and Interaction Components

Date	29/03/2019
Author(s)	Yuchen Zhao, Elena Simperl
Dissemination level	Public
Work package	WP4: Interaction Design and Storytelling
Version	2.0

Document metadata

Quality assurers and contributors

Quality assurer(s)	Ahmet Soylu, Till C. Lech (SINTEF)
Contributor(s)	

Version history

Date	Version	Description
08/03/19	0.1	Initial draft for internal review
23/03/19	0.2	References added
27/03/19	0.3	Section 4 revised
31/03/19	Final	Final version for submission

Executive summary

This document presents Deliverable 4.4 “Visualisation and Interaction Components v2” of the TheyBuyForYou (TBFY) project. It contains a brief survey about the usage of data visualisation in existing public procurement data portals, a complete guideline on data visualisation and interaction design, and a preliminary user study plan.

This deliverable is part of WP4 “Interaction Design and Storytelling”, which contributes to the TBFY project by creating an automatic data storytelling tool, increasing engagement with the TBFY knowledge graph through novel interfaces, and helping the business cases with any user-centric aspects.

This deliverable aims to:

- Help the project partners understand how data visualisations are used in the real world.
- Provide information about how data visualisations should be designed in theory, which helps with the design of the interface of the TBFY knowledge graph.
- Propose a number of user studies that help the partners improved their business cases.

Table of Contents

DOCUMENT METADATA	2
EXECUTIVE SUMMARY	3
1 INTRODUCTION	6
2 DATA VISUALISATION IN EXISTING PROCUREMENT DATA PORTALS	6
2.1 RELATED PROCUREMENT DATA PORTALS	6
2.2 VISUALISATION INDEXED BY ANALYSIS.....	10
3 DATA VISUALISATION DESIGN IN THEORY: A COMPLETE GUIDELINE	14
3.1 PRACTICAL GUIDELINES – DASHBOARDS	15
3.1.1 <i>What are dashboards</i>	15
3.1.2 <i>What do dashboards display</i>	16
3.1.3 <i>What do dashboards look like</i>	16
3.1.4 <i>How to design a dashboard</i>	18
3.1.5 <i>How to improve a dashboard</i>	18
3.2 PRACTICAL GUIDELINES – CHARTS.....	19
3.2.1 <i>Example analysis 1: how variables change over time</i>	19
3.2.2 <i>Example analysis 2: parts of a whole</i>	20
3.2.3 <i>Example analysis 3: comparing values in different categories</i>	21
3.2.4 <i>Example analysis 4: showing values on a map</i>	22
3.2.5 <i>Example analysis 5: comparing procurement indicators</i>	24
3.2.6 <i>Example analysis 6: displaying relationships between items</i>	24
3.2.7 <i>Example analysis 7: displaying processes, tasks, and other time-based activities</i>	25
3.2.8 <i>Example analysis 8: showing statistics and distributions</i>	25
3.2.9 <i>Example analysis 9: communicating combined analyses</i>	26
3.3 GENERAL GUIDELINES.....	26
3.3.1 <i>Things to do</i>	26
3.3.2 <i>Things NOT to do</i>	29
4 WP4 SUPPORT FOR PILOTS AND DATA STORIES	31
4.1 COMMON TASKS OF PROCUREMENT DATA PORTAL USERS.....	31
4.2 USER STUDIES	32
4.3 DATA STORIES FOR OPENOPPS	33
5 CONCLUSION	34
6 REFERENCES	35

Table of Figures

Figure 1. A sample dashboard for marketing analysis (Figure from [1]).	16
Figure 2. Different areas in a dashboard. The top-left area and the central area are emphasised and viewers will pay more attention to them. The top-right and bottom-left areas have lower emphasis. The bottom-right area has the lowest emphasis. (Figure from [1])	17
Figure 3. An example of a line chart. (Figure from https://datavizproject.com/)	20
Figure 4. An example of a pie chart. (Figure from https://datavizproject.com/)	20
Figure 5. An example of a treemap (Figure from https://datavizproject.com/)	21
Figure 6. An example of a bar chart. (Figure from https://datavizproject.com/)	22
Figure 7. An example of a stacked bar chart. (https://datavizproject.com/)	22
Figure 8. An example of a choropleth map. (Figure from https://datavizproject.com/)	23
Figure 9. An example of a bar chart on a map. (https://datavizproject.com/)	23
Figure 10. An example of a connection map. (https://datavizproject.com/)	23
Figure 11. An example of a radar chart. (https://datavizproject.com/)	24
Figure 12. An example of network graph. (https://datavizproject.com/)	24
Figure 13. An example of a Gantt chart. (https://datavizproject.com/)	25
Figure 14. An example of a violin chart. (https://datavizproject.com/)	26
Figure 15. An example of column sparklines. (https://datavizproject.com/)	26
Figure 16. People consider that the bars close to each other belong to the same group, even if they have different colours.	28
Figure 17. Items with the same colour (left: black and white), the same size (middle: big and small), or the same shape (right: square and triangle) are perceived to belong to the same group.	29
Figure 18. Example of a pie chart with many categories. (Figure from http://www.storytellingwithdata.com/blog/2011/07/death-to-pie-charts)	29
Figure 19. Example of a misleading 3D chart. (Figure from https://en.wikipedia.org/wiki/Misleading_graph)	30
Figure 20. Example of a heatmap using rainbow colourmap. (Figure from [3])	30
Figure 21. Example of a line chart using double y-axes. (Figure from https://www.businessinsider.com/travel-agents-vs-online-hotel-bookings-2015-12?r=UK)	31
Figure 22. Data story about Brexit's influence on public procurement (https://public.tableau.com/profile/yuchen.zhao#!/vizhome/BrexitImpactonUKProcurement/Dashboard1?publish=yes)	33

1 Introduction

Visualisations are widely used to communicate complex data, information, and knowledge in an intuitive way. Powered by the knowledge graph of the TBFY project, data visualisations can be used to reveal important trends and patterns in public procurement, to inform business decisions, and to improve the user experience of many business cases. In this deliverable, we present the second version of the visualisation and interaction components of WP4.

The aim of this deliverable is to provide help in terms of data visualisation and interaction from different aspects. We first investigate existing data visualisation usage in public procurement to help the TBFY project partners choose appropriate visualisation schemes based on their specific business cases and customer demands. Second, we provide a complete guideline on data visualisation design from established theories and practices. This guideline will help our partners review and improve their current visualisation design. We finally propose a number of user studies that would help evaluate the effectiveness of the visualisations used in the partners' business cases.

This deliverable can help audience both within and outside the TBFY project. The project partners can use this deliverable to find the gap between the analyses that their products have supported and those supported by the existing procurement data portals. In addition, they can use the data visualisation and interaction guideline to review their current product design. Outside the TBFY consortium, the guideline can also be used as general instructions for dashboard design and for visualisation optimisation.

There are several other deliverables related to this document. The business cases described in D6.1 can benefit from the Section 2 and 3 in this document. The interface of the TBFY knowledge graph, which will be delivered in D2.2, will be designed based on the guideline in this document. The data visualisation design of D4.3, which belongs to WP4, will also be based on the guideline.

The rest of this deliverable is organised as follows. Section 2 reports our survey about 28 existing procurement data portals. It focuses on what kind of analyses these portals have provided and what kind of visualisations they have used to support the analyses. Section 3 is the data visualisation design guideline. It focuses on both practical guidelines for specific examples and general guidelines in visualisation design. Section 4 presents how WP4 supports partners through pilots of user studies to evaluate data visualisations and making data stories from public procurement data.

2 Data Visualisation in existing procurement data portals

In this section, we first briefly introduce the data portals used in our survey. Then we summarise the analyses provided by these portals, and index the used visualisations by the analyses. The aim of this section is to help the partners find the analyses of their interests in their business cases and understand what kind of visualisations they can use to support the analyses.

2.1 Related procurement data portals

As shown in Table 1, there are 28 different data portals in our survey. We started our search from the website of Open Contracting Partnership (<https://www.open-contracting.org/>) and then iterated and expanded our search based on the results we had. These data portals are from different countries and

are related to public procurement or public spending. Detailed information can be found through the links provided with these samples.

Name	Link	Description
Tendertracking	http://www.tendertracking.eu/	Launched by Corruption Research Center Budapest (CRCB). Users are able to access more than 100,000 public procurement contracts.
Opentender	https://opentender.eu/	Part of the DIGIWHIST project (http://digiwhist.eu/). Users can access, download, and analyse data from different sources.
Openbudgets	http://openbudgets.eu	Users can upload, visualise, and analyse public budget and spending data through Openbudgets.
Yourdatastories	http://platform.yourdatastories.eu/select-dashboard	Aggregating contract data from different sources and providing dashboards for analysis. The three dashboards are for the analysis of Greek public contracts, the analysis of European contracts, and the analysis of country comparison.
zIndex	https://www.zindex.cz/	Analysing the contract data from the Public Procurement Notice (https://www.vestnikverejnychzakazek.cz/) and calculating zIndex scores for different contracting authorities.
Vue sur les contrats	https://ville.montreal.qc.ca/vuesurlescontrats/	Analysing the contract data from the City of Montreal's open data

		portal (http://donnees.ville.montreal.qc.ca/) and providing visualisations.
Clearspending	https://clearspending.ru/	Analysing the contract data from the Russian Unified Information System (http://zakupki.gov.ru/) and allowing users to search contracts, budgets, buyers, and sellers.
Redflags	https://www.redflags.eu/	Analysing the procurement data from Tenders Electronic Daily (TED) and using algorithms to “red flag” risky procurements.
Development Gateway - Monitoring and Evaluation (M&E) Dashboard	http://ocdemo.developmentgateway.org/ui/index.html	Analyzing, monitoring, and evaluating information on public procurement specifically designed to help users understand procurement efficiency, and the competitiveness and cost-effectiveness of public markets.
EDCA Visualisation - Nuevo Aeropuerto	https://datos.gob.mx/nuevoaeropuerto/	Visualizations and resumes about the Nuevo Aeropuerto Internacional de la Ciudad de México contracts.
Moldova Open Contracting Portal	http://opencontracting.date.gov.md/	Visualisations of Moldova contracts.
Development Gateway - Corruption Risk Dashboard	http://ocdemo.developmentgateway.org/ui/index.html#!/crd	Using a red flagging approach to help users understand the potential presence of fraud, collusion or rigging in public contracting.

Testigo Social 2.0 / Social Witness 2.0	https://testigosocial.mx/	Allowing citizens to know how and where government procures.
DNCP Visualizaciones	https://www.contrataciones.gov.py/datos/visualizaciones	Providing multiple contract visualisations: including quantity + amount, stages of each contracting process with dates and delivery schedule.
ProZorro Business Intelligence	http://bi.prozorro.org/sense/app/fba3f2f2-cf55-40a0-a79f-b74f5ce947c2/sheet/HbXjQep/state/analysis/select/Language/EN#view/pEh	Performing analysis of procurement in a chosen area: goods, categories, tenders, lots, tenderers and complaints in Ukraine procurement. It is built on top of the proprietary QlikSense platform.
Quiencobra laobra	http://quiencobralaobra.es/	Containing all public works contracts published in the Spanish Official Gazette between 2009 and 2015
Budeshi	http://www.budeshi.ng/home/data	Linking budget and procurement data to various public services. It is accessible to the public to interact with and make their own comparisons.
Mexico's Federal Government Open Contracting Portal	https://www.gob.mx/contratacionesabiertas/home#!//	Providing information on all stages of the contracting process, specifically designed to link with key context data such as public infrastructure, service delivery and public spending data.
Mexico Procurement	http://mexico.procurement-analytics.org/	Providing insights into Mexico's procurement performance.

nt Dashboard		
Kosovo: e-prokurimi	http://e-prokurimi.org	Visualization of procurement data for the municipalities of Kosovo.
CDMX Portal	http://www.contratosabiertos.cdmx.gob.mx/contratos	Providing access to data on contracts in Mexico City.
PPTIN	http://ppip.gov.np/	An open contracting portal for Nepal
Dozorro Risk Indicators	http://risk.dozorro.org/	Showing risk indicators for tenders in Ukraine taking place through the ProZorro platform
Vue sur les contrats (FFunction)	https://ffctn.com/a/contrats/	Enabling citizens to understand how municipal funds are used
BA Obras	http://www.buenosaires.gob.ar/baobras	
Zambia OCDS Reports	https://www.zppa.org.zm/ocds-reports	Procurement analytic reports driven by OCDS data.
USASpending	https://www.usaspending.gov	Tracking federal spending to ensure taxpayers can see how their money is being used in communities across America.
Bivol	https://dfz.bivol.bg/	Developing databases and search engines to help investigative journalists, civil activists and public transparency

Table 1. Related data portals

2.2 Visualisation indexed by analysis

We now summarise what kind of analyses have been provided by the portals in our survey. For each of the analyses, we show the types of visualisations have been used by the portals to support it.

- Total number (total volume) of tenders (contracts / bids) over time

- Bar chart
 - Tendertracking.eu
 - Opentender.eu
 - Quiencobraobra.es
 - Budeshi
- Stacked bar chart
 - M&E dashboard
- Line chart
 - M&E dashboard
 - Moldova open contracting portal
 - Mexico Procurement Dashboard
 - Kosovo: e-prokurimi
 - PPTIN (Public Procurement Transparency Initiative in Nepal)
- Pie chart
 - Testigo Social 2.0
- Heat map
 - DNCP Visualizaciones
- Pareto Chart (line chart + bar chart)
 - ProZorro Business Intelligence
- Total number (total volume) of tenders (contracts) by sector (contract type) over time
 - Stacked bar chart
 - Vue sur les contrats
 - Grouped bar chart
 - Redflags
- Total number of flagged tenders (contracts) over time
 - Line chart
 - Redflags
 - Corruption risk dashboard
- Number of registered sellers over time
 - Line chart
 - PPTIN (Public Procurement Transparency Initiative in Nepal)
- Number (volume) of tenders (contracts) of each buyer (seller) overtime
 - Line chart
 - Yourdatastories
 - Bubble chart
 - Yourdatastories
 - Nuevo Aeropuerto
 - DCNP Visualizaciones
- Average indicator scores over time (an indicator is a metric designed to measure a specific aspect of a tender)
 - Bar chart (without error bar)
 - Tendertracking.eu
 - Opentender.eu
 - M&E dashboard

- Stacked bar chart (without error bar)
 - M&E dashboard
- Line chart (without error bar)
 - M&E dashboard
- Average indicator scores by sector (top tender)
 - Bar chart (without error bar)
 - Opentender.eu
 - Pareto Chart (line char + bar chart)
 - Dozorro Risk Indicators
- Average volume of tenders (contract) over time
 - Bar chart (without error bars)
 - Tendertracking.eu
 - Opentender.eu
- % of tenders within a certain indicator score range over time
 - Bar chart
 - Opentender.eu
- Distribution of the number of indicators (flags)
 - Pie chart
 - Redflags
 - Donut chart
 - Redflags
- Buyers' (Sellers') offices, tenders' locations
 - Map
 - Tendertracking.eu
 - DNCP Visualizations
- Data of individual regions
 - Map
 - Opentender.eu
 - Yourdatastories
- Number of buyers (sellers) by region
 - Choropleth map
 - Opentender.eu
- Total number (volume) of tenders (contracts) (per buyer / per seller) by region
 - Map
 - M&E dashboard
 - Bivol
 - Choropleth map
 - Opentender.eu
 - Yourdatastories
 - ProZorro Business Intelligence
 - Kosovo: e-prokurimi
 - Zambia OCDS Reports
 - Bivol
 - Bar chart

- Budeshi
- Budgets by functional classification (chief controller / special-purpose item / type of budget spending code / agency / object class)
 - Tree map
 - Clearspending
 - USASpending
 - Bubble tree
 - Clearspending
- Number (Volume / %) of tenders (contracts, bids) by sector (procurement type / procedure type / tender status / top buyers / top sellers / top seller groups / buyer country / seller country / seller size / CPV / number of bidders / volume of contract)
 - Tree map
 - Opentender.eu
 - Yourdatastories
 - Quiencobralaobra.es
 - Kosovo: e-prokurimi
 - Vue sur les contrats (FFunction)
 - Bar chart
 - Opentender.eu
 - Yourdatastories
 - zIndex
 - M&E dashboard
 - Moldova open contracting portal
 - Budeshi
 - Dozorro Risk Indicators
 - Zambia OCDS Reports
 - Grouped bar chart
 - Dozorro Risk Indicators
 - Zambia OCDS Reports
 - Pie chart
 - Yourdatastories
 - Testigo Social 2.0
 - ProZorro Business Intelligence
 - Kosovo: e-prokurimi
 - Zambia OCDS Reports
 - Donut chart
 - zIndex
 - Nuevo Aeropuerto
 - Stacked bar chart
 - ProZorro Business Intelligence
 - Quiencobralaobra.es
 - Polar area chart
 - Mexico's Federal Government Open Contracting Portal
 - Line chart
 - Zambia OCDS Reports

- Average saving by procurement type (number of bidders)
 - Line chart
 - Zambia OCDS Reports
 - Bar chart
 - Zambia OCDS Reports
- Distribution of tenders (contracts / sellers / top buyers)
 - Packed circle chart
 - Nuevo Aeropuerto
 - CDMX Portal
 - PPTIN (Public Procurement Transparency Initiative in Nepal)
- Distribution of volume of tenders (contracts)
 - Histogram
 - Mexico Procurement Dashboard
 - Box plot
 - Mexico Procurement Dashboard
- Correlation between the amount spend by a buyer and the number of different sellers that contract with that buyer
 - Bubble chart
 - Mexico Procurement Dashboard
- Indicator scores
 - Polar area chart
 - Opentender.eu
 - Colour map
 - Opentender.eu
 - Solid Gauge chart
 - Opentender.eu
 - zIndex
- Phases of call for bids
 - Tree
 - DNCP Visualizaciones
- Number of days by procurement phases
 - Stacked bar chart
 - Mexico Procurement Dashboard
- Number (volume) of risky tenders in the top buyers
 - Heat map
 - Dozorro Risk Indicators
- Average number of days taken to complete a procurement process by procuring entity
 - Line chart
 - Zambia OCDS Reports

3 Data Visualisation design in theory: A complete guideline

In Section 2, we discussed how data visualisations are used in the products in the real-world. In this section, we focus on the theoretical aspect of data visualisation design, which can help the partners

review their business cases. We provide guidelines for organisations that publish or work with procurement data on visual and interaction design. In some of the project partners' business cases, they will process different types of procurement data such as data about tenders, buyers, sellers, and contracts. In the data, there will also be different types of variables including numerical (e.g. volume of money), temporal (e.g. time), categorical (e.g. sector of spending), and spatial (e.g. coordinates/polygons). To achieve the goals (e.g. analysis of spending and budget data for BMBV, facilitating the understanding of information for EDP, and facilitating transparency for SEPPAS), the data needs to be presented to users in a meaningful, effective, and concise way. Therefore, we provide a collection of guidelines on how to use data visualisation to fulfil different requirements in the business cases.

The guidelines contain different aspects of data visualisation, including dashboard design, chart recommendations for different tasks, as well as good / bad practices. This aims to help the partners make design decisions based on their needs and also to provide suggestions for reviewing the current designs.

The rest of this section is organised in three parts. We first start with practical guidelines about what analytical functions a data portal should have and how to arrange and organise these functions in an integrated manner as a dashboard. We then introduce practical guidelines for choosing individual charts according to the different requirements of the business cases. Finally, we discuss general guidelines that apply to both dashboards and individual charts about what should be considered and what should be avoided in data visualisation design.

3.1 Practical guidelines – Dashboards

Users of a procurement data portal should be able to analyse all the data of their interest in one place. The presentation of the data should be well organised and in an accessible style. We introduce the analysis that a procurement data portal may aim to support and how to integrate the presentation of such analysis through a dashboard. This sub-section is based on the book about dashboard design by Few [1].

3.1.1 What are dashboards

According to the definition by Stephen Few [1]:

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance."

In the definition, there are several requirements for designing a dashboard

- A dashboard has to be a visual display
- The information shown in a dashboard is needed to achieve specific goals
- A dashboard has to fit on a single screen
- The information shown in a dashboard should be able to be comprehended at a glance

Fig. 1 is a sample dashboard for marketing analysis. It uses different visualisations including line charts, bar charts, and sparklines to communicate different information.

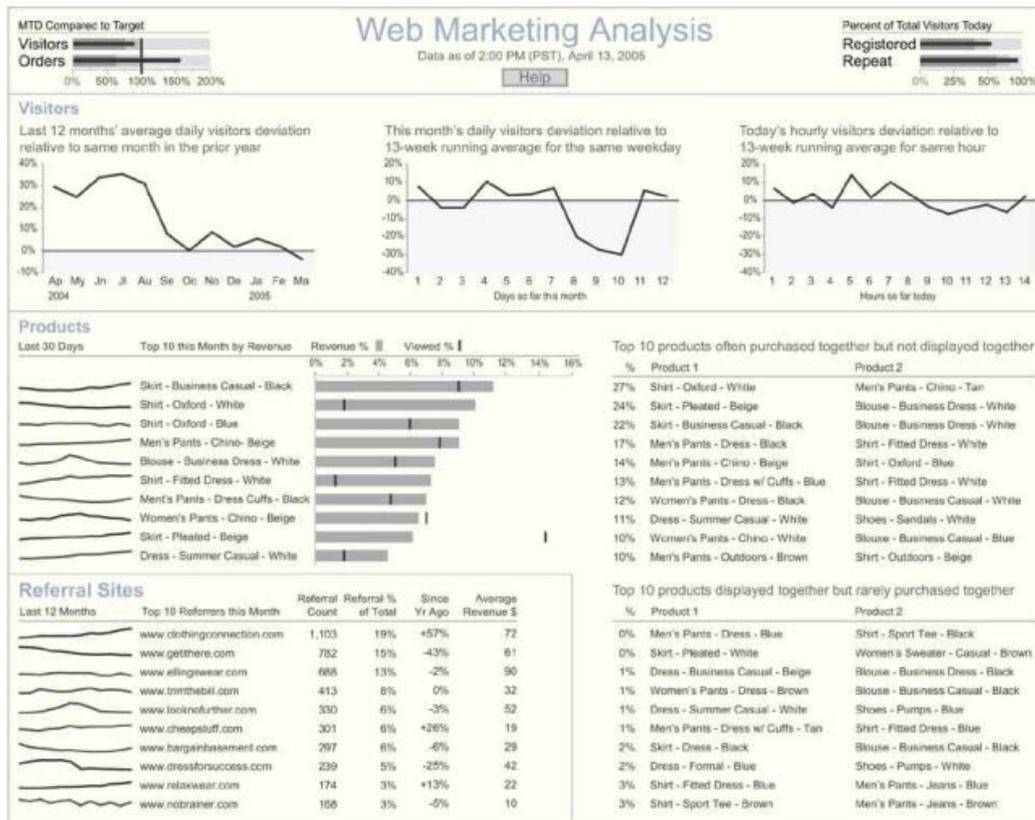


Figure 1. A sample dashboard for marketing analysis (Figure from [1]).

3.1.2 What do dashboards display

In a dashboard, there are some commonly used analyses presented through visualisation. For a procurement data portal, its dashboard may contain the following analyses:

- Analysis over time: such as contracts during a previous period, contracts to date, and upcoming opportunities.
- Comparison: such as comparison between buyers, sellers, sectors, and regions.
- Evaluation: such as key procurement-related indicators, contract completeness, and uncertainty.
- Structure: such as composition of a whole, relationships, and geographical distribution.

3.1.3 What do dashboards look like

As a dashboard uses one screen to show multiple contents at the same time, it is important to arrange these in an effective way to let the viewers focus on the information that is most important to them.

A dashboard can be divided into five different areas, i.e. central, top-left, top-right, bottom-left, and bottom-right. These areas have different levels of emphasis and viewers pay more attention to the contents in the areas with high emphasis. The levels of emphasis in different areas of a dashboard are:

- High emphasis: Top-left area, central area

- Medium emphasis: Top-right area, bottom-left area
- Low emphasis: Bottom-right area

Dependent on the goal of the dashboard, important contents should be put in the top-left or the central area. For example, in the sample marketing analysis dashboard shown in Fig. 2, the most important information is how many visitors have visited the web in the last year and which products contribute mostly to revenue. So the line chart for the visitor change in the last 12 months and the bar chart of the top-10 products for revenue are shown in the emphasised areas. Less important information such as daily visitor change, monthly revenue change, and referral sites is shown in the top-right area and the bottom-left area, which have medium emphasis. The least important information such as the least displayed products and the least purchased products is shown as tables in the bottom-right area.

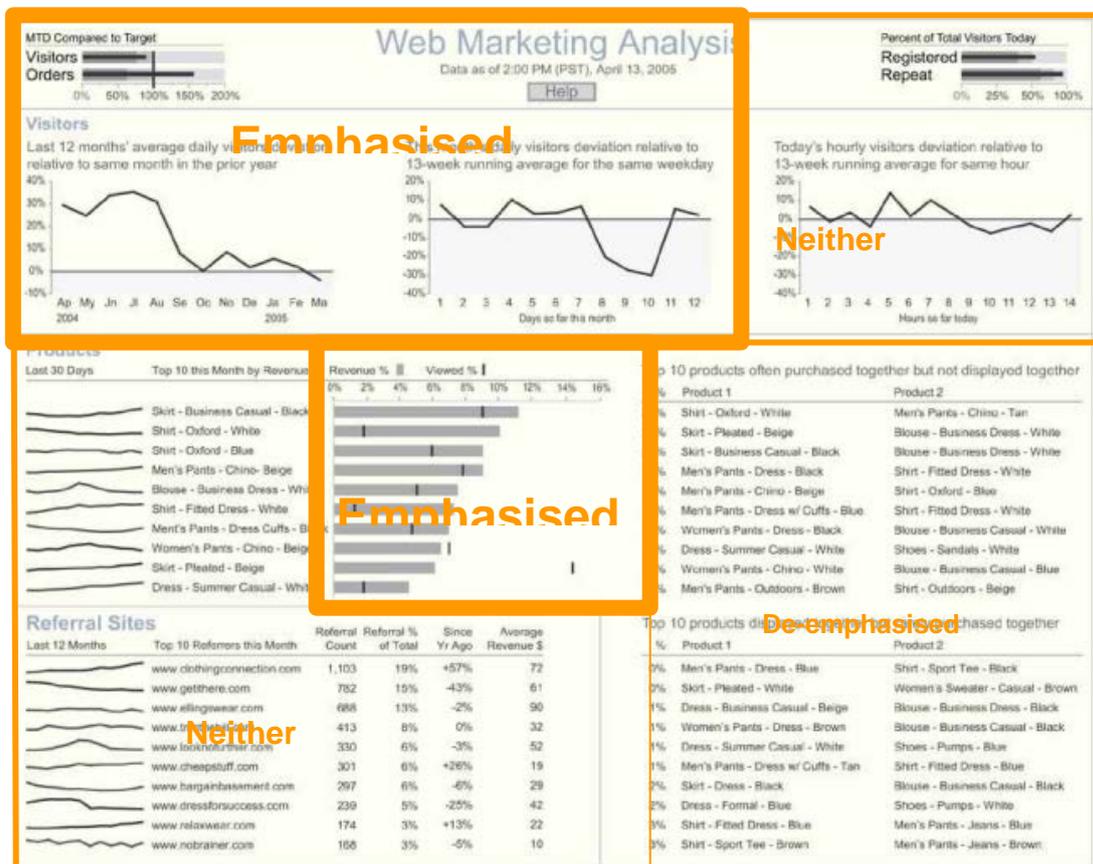


Figure 2. Different areas in a dashboard. The top-left area and the central area are emphasised and viewers will pay more attention to them. The top-right and bottom-left areas have lower emphasis. The bottom-right area has the lowest emphasis. (Figure from [1])

There are also other aspects that should be considered when organising the information in a dashboard:

- The information in a dashboard should be divided into groups. There are different ways of grouping information:

- Business functions: for example, for a procurement data portal, these functions could be tendering, bidding, performance of procurement, etc.
- Entities: for example, the sample dashboard in Fig. 2 uses visitors, products, and referral sites. For a procurement data portal, it can use entities such as buyers, sellers, regions.
- Uses of the data: such as comparison, correlation.
- Once the groups have been decided, the items that belong to the same group should be co-located close to one another in the dashboard.
- Different groups should be delineated using the least visible means. If the dashboard has enough space, then use white space to delineate the groups, otherwise use subtle borders (light lines).

3.1.4 How to design a dashboard

- Select the most important information for the viewer of the dashboard
- For each of the selected items, decide:
 - level of summarisation for the item
 - unit of measure
 - contextual information to be added
 - charts that can show the information most effectively
 - importance of the item to the viewer
 - location of the item in the viewing sequence
 - other items that the user may want to compare with this item
- Group information and arrange items according to the groups
- Reduce the components that do not communicate any data
- Enhance the components that communicate important information

3.1.5 How to improve a dashboard

For the design of a dashboard, there are several pitfalls that should be avoided:

- Exceeding the boundaries of a single screen: for example, requiring scrolling to view a dashboard or dividing a dashboard into separate screens makes the dashboard difficult to view.
- Supplying inadequate context for the data in the dashboard: simply showing a numbers or a chart sometimes cannot clearly explain what is going on. What does the number mean? Is it good or bad and compared to what? What does the chart try to say? Failing to provide context to the data makes the data less meaningful. In the Visitors group of the sample dashboard in Fig. 2, each line chart has a short description that adds context to help the viewers to understand the data.

- Displaying excessive detail or precision: showing variables with extremely high precision is unnecessary. For example, do we need to show the level of seconds of a time field? Do we need to show long decimal digits of numbers?
- Using a deficient measure: for example, to show the difference in the comparison of two volumes (money), using the absolute difference such as “£20,000,000 more than” is less efficient than using percentage “10% higher than”.
- Using inappropriate, poorly designed, or meaningless variety of charts: some charts are only good at certain tasks. Using inappropriate charts may fail to communicate or distort the data. For detailed suggestions about charts for specific tasks, check Section 3.
- Encoding data inaccurately: for example, encoding the start point of the y-axis of a bar chart with values other than 0 may distort the results and lead to incorrect conclusions.
- Poor layout and arrangement: for example, using important areas (top-left and central) of a dashboard to show unimportant information (logos, navigation bars), or showing items that belong to the same group separately.
- Lack of highlighting important data: for example, giving same visual importance (e.g. size, colour, boldness) to all the items will make the viewers fail to find out which items they should focus on. One solution is to change the type of presentation. For example, in the sample dashboard in Fig. 2, the information in the bottom-right area is shown as tables rather than charts, to de-emphasise its importance.
- Cluttering with useless visual components: using too many decorations that do not communicate any information from the data will distract the viewers
- Misusing or overusing colours: using colours without consideration can be misleading. Viewers are likely to pay attentions to colour contrast. They also tend to associate objects with same colours together (see section 4.1.9.). If two objects in a line chart and a bar chart do not represent the same thing, then they should not be encoded with the same colour.
- Unattractive: dashboards should be aesthetically pleasing. Carefully think about aspects such as colours, resolution, and fonts.

3.2 Practical guidelines – Charts

In section 2, we have introduced how to design a dashboard. In this section, we discuss what types of charts are suitable for the different analysis in a dashboard. The analyses are summarised from a number of existing public procurement data portals and the requirements provided in D6.1. For each one of them, we discuss what it is, what kind of questions in procurement it can answer, and what kind of charts can be used to support the task.

3.2.1 Example analysis 1: how variables change over time

The change of key variables such as the volume (money) of contracts and the number of contracts is important information. Visualising such information can help reveal trends in the data. Such visualisations can answer questions such as:

- Does this increasing trend mean more opportunities in the market?

- Is the local council increasing its investment in a certain sector in recent years?

Line charts (as shown in Fig. 3) are one of the most commonly used techniques to visualise changes over time. The x-axis of a line chart represents time and the y-axis represents the analysed dependent variable. The value of the variable at each time point is presented as a dot and adjacent dots are connected with lines. Multiple lines can be shown at the same time in one line chart and can be distinguished through colours. A rule of thumb is that the number of lines in one chart should not be more than 5-7.

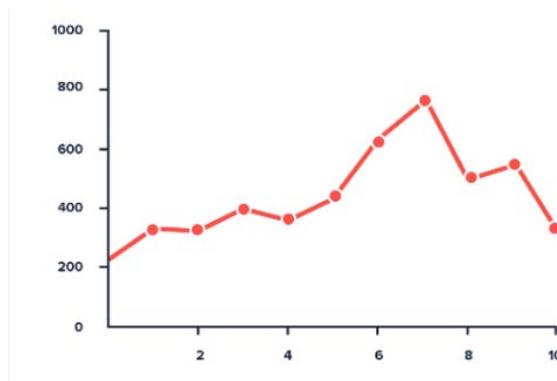


Figure 3. An example of a line chart. (Figure from <https://datavizproject.com/>)

3.2.2 Example analysis 2: parts of a whole

Composition visualisation can help people understand the components and structure of a subset of the data. For example, people may want to know:

- How does my local council spend the public money in different types of projects?
- Who are the contract winners and how much money did they get from the buyer?

Pie charts (as shown in Fig. 4) can be used to show parts of a whole. A pie chart is divided into separate sectors (like a sliced pie) and each sector represents a category. The angle of a sector is proportional to the value of the presented category. Different sectors are distinguished by colours or annotations. But when the number of categories in a pie chart is more than 4, it becomes hard to read.

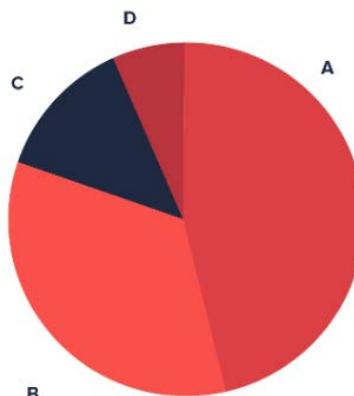


Figure 4. An example of a pie chart. (Figure from <https://datavizproject.com/>)

Treemaps (as shown in Fig. 5) are commonly used by many portals to show compositions of contracts (number and volume of money), buyers, sellers, and so on. A treemap is a rectangular area divided into small rectangular sub-areas. The size (area) of each rectangle represents the value of the corresponding category. As a treemap has a nested structure, it can also show the hierarchical structure in the data. Compared with pie charts, treemaps can represent more categories without reducing the readability.

It is worth to note that the values of categories are encoded into area sizes. This makes it difficult for people to directly compare different components in a treemap. According to the theory of Cleveland and McGill [2, 5], people are not good at comparing areas.

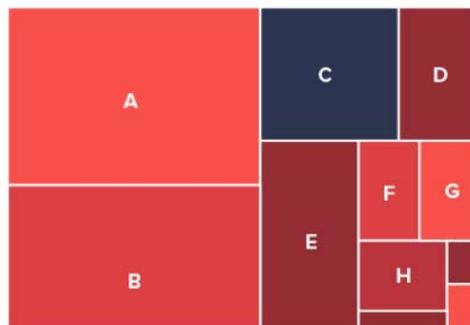


Figure 5. An example of a treemap (Figure from <https://datavizproject.com/>)

3.2.3 Example analysis 3: comparing values in different categories

Comparison is one of the most important tasks in data analysis. Through comparison, we can distinguish between high and low, good and bad, and so on. Many questions can be answered by visualising comparisons, such as:

- Who are the top 10 contract winners and what is the difference between their contract volumes?
- What is the difference in the number of awarded contracts between big companies and small businesses?

Bar charts (as shown in Fig. 6) are commonly used techniques for comparison between multiple categories. The x-axis represents all the categories and the y-axis represents the values of the categories. It is easy for people to compare the values of two categories in a bar chart, as their values are encoded as the heights of bars and people are good at comparing positions along a common scale.

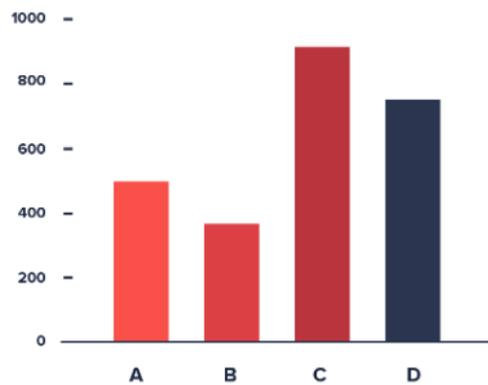


Figure 6. An example of a bar chart. (Figure from <https://datavizproject.com/>)

Stacked bar charts (as shown in Fig. 7), as a variety of bar charts, can present both comparison and the composition of a whole. For each bar of a category, it can be divided into several sub-categories. The heights of the sub-categories are proportional to their sizes in the whole category. However, the comparison between different sub-categories may be less accurate than in a standard bar chart.

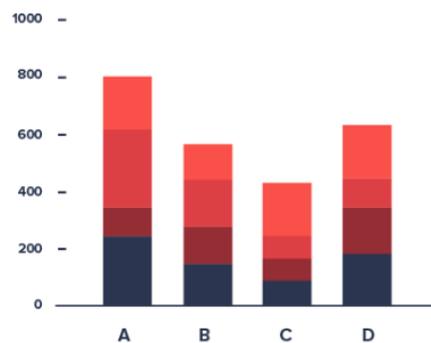


Figure 7. An example of a stacked bar chart. (<https://datavizproject.com/>)

3.2.4 Example analysis 4: showing values on a map

As procurement data commonly comes with geography related information, visualising dependent variables based on such information can provide insights about differences between regions (e.g. cities, counties, countries, and continents). It can help answer questions such as:

- Does a particular buyer only spend money on domestic sellers, rather than international sellers?
- What is the difference between the transparency scores of different European countries?

Choropleth maps (as shown in Fig. 8) are maps whose different areas are shaded or coloured. The colour saturation or hue of an area is proportional to the value of the variable in the area. It can provide an overview about how the variable is distributed geographically. But as people's accuracy on comparing colour saturation or hue is low, using choropleth maps for precise comparison is not suggested.

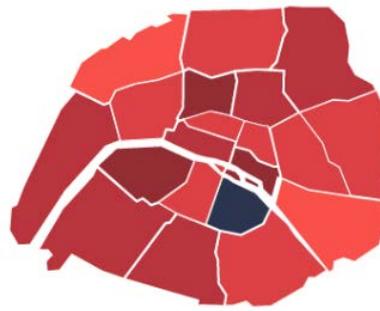


Figure 8. An example of a choropleth map. (Figure from <https://datavizproject.com/>)

Showing bar charts on a map (as shown in Fig. 9) can provide more accurate comparison than using a choropleth map. For each area, the height of the bar represents the value in that area.

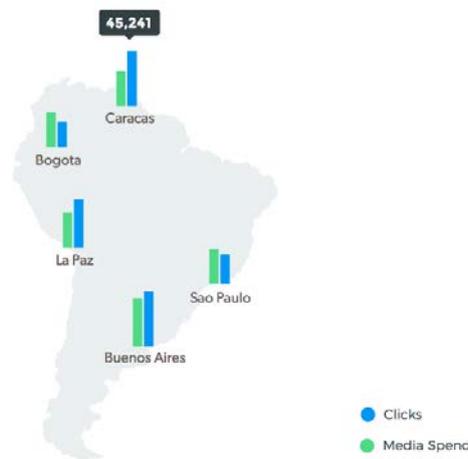


Figure 9. An example of a bar chart on a map. (<https://datavizproject.com/>)

Connection maps (as shown in Fig. 10) can show relationships between different areas of a map. Each area on the map is marked by a circle and the relationships (e.g. trade, transport) between different areas are presented as lines between the circles. The width of the lines can also be adjusted to show different levels of strength of relationships.



Figure 10. An example of a connection map. (<https://datavizproject.com/>)

3.2.5 Example analysis 5: comparing procurement indicators

Some procurement data portals use many procurement-related indicators to show the transparency, integrity, and risks of a procurement process. Each of these indicators can only describe one aspect. Visualising them in an integrated way may help people understand the overall procurement performance of a buyer, a seller, or a region.

Radar charts (as shown in Fig. 11) are commonly used to present multiple variables related to the same entity. For a radar chart, there are several spokes that start from the centre. These spokes divide the space into equal sized sectors and each of them represents a variable. The data length of a spoke is proportional to the value of the variable and adjacent data values are connected with straight lines.

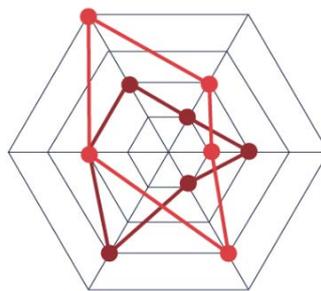


Figure 11. An example of a radar chart. (<https://datavizproject.com/>)

3.2.6 Example analysis 6: displaying relationships between items

One important requirement for a procurement audit is to reveal potential collusion and corruption. To achieve this, we need to know the relationships between buyers and sellers. Visualising such relationships can help answer questions such as:

- Is there a small cluster of sellers that always win the contracts from a buyer?

Network graphs (as shown in Fig. 12) can present the entities of a network and the relationships between these entities. Each entity is presented as a node and the relationship between two nodes is presented as an edge. Through a network graph, we can find out who is in the centre of the network and whether there are clusters in it.

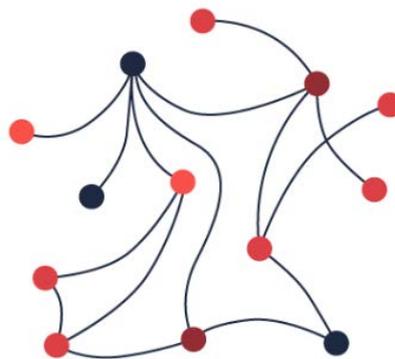


Figure 12. An example of network graph. (<https://datavizproject.com/>)

3.2.7 Example analysis 7: displaying processes, tasks, and other time-based activities

There are several stages in the life cycle of a contract in public procurement. Sellers may be interested in knowing what the upcoming bidding opportunities are and which contracts are going to end. Visualising such time based events can help sellers prepare their future bidding strategies.

Gantt charts (as shown in Fig. 13) can visualise multiple time-based events in a juxtapositional style. The x-axis of a Gantt chart represents the timeline and the y-axis represents different events. Each event is shown as a bar. The left end and the right end of a bar represent the start and end time of the event, respectively.

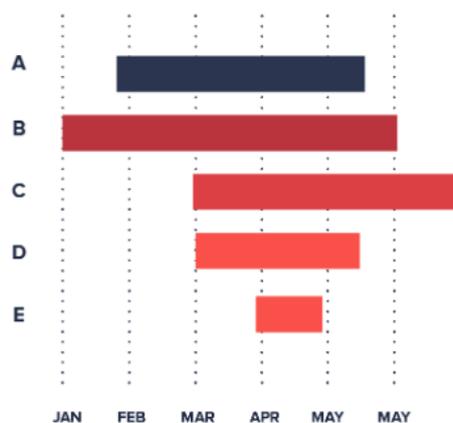


Figure 13. An example of a Gantt chart. (<https://datavizproject.com/>)

3.2.8 Example analysis 8: showing statistics and distributions

Basic statistics such as average, median, maximum, and minimum are often provided in data analysis. These results are highly summarised, which means that they may not reveal the whole picture of the data. Even for similar statistical results in a procurement data set, people may want to have deeper insights such as:

- What is the difference in the distributions of the contract volumes of these two sellers who have similar average contract volumes?
- Is the distribution of the contract volumes of this buyer more “uniform” or more “long tail”?

Violin charts (as shown in Fig. 14) can show the distribution of numeric data. A violin chart is the combination of a box plot [6] (i.e. a plot showing median, higher/lower quartiles, maximum/minimum, and outliers) and the probability density of the data. It can present both the important statistics of the data and the distribution.

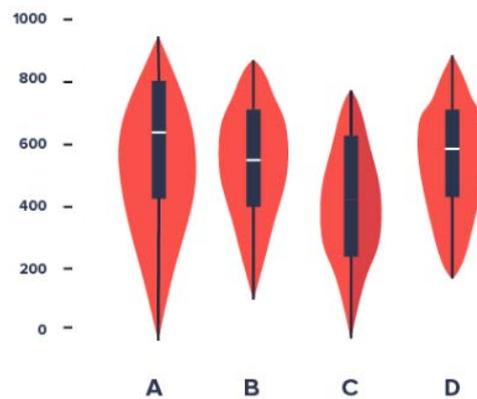


Figure 14. An example of a violin chart. (<https://datavizproject.com/>)

3.2.9 Example analysis 9: communicating combined analyses

Visualisation may tell better stories when combined with other components such as texts and tables. An ideal analytical report should combine these components effectively.

Column sparklines (as shown in Fig. 15) are word-sized visualisations that can be positioned anywhere in a document, a table, or a dashboard. A column sparkline uses bars to present values and can be used to show trends over time or to show comparison. As it is word-sized, it is easy to use it in a sentence to explain a complex trend. Multiple sparklines can be used in a repeated and intensive way (e.g. in a column of a table, or like the sample dashboard shown in Fig. 1) to communicate more information than by using just numbers.



Figure 15. An example of column sparklines. (<https://datavizproject.com/>)

3.3 General guidelines

We now provide some general design guidelines about good and bad practices in data visualisation. Many of these general guidelines apply to both dashboards and individual charts.

3.3.1 Things to do

3.3.1.1 Choose high accuracy charts for comparison

As discussed above, comparison is an important task in data analysis. There are several charts that can be used for comparison (e.g. bar charts, tree maps, and pie charts), but the expected levels of accuracy are different.

According to the theory by Cleveland and McGill [2, 5], people have different accuracy when comparing different visual elements. The ranks of accuracy (from high to low) are:

- Position
- Length
- Angle
- Slope
- Area
- Volume
- Density
- Colour

For example, using bar charts to visualise data can help people make more accurate comparison than using pie charts. Because in bar charts people compare positions and it has higher rank of accuracy than comparing angles in pie charts. Similarly, bar charts are also more accurate than heat maps, as comparing position is more accurate than comparing colours. Therefore, if you want the viewers to compare values accurately through visualisation, you should give preferences to the charts with high accuracy, e.g. bar charts and stacked bar charts (comparing positions and lengths), rather than charts with low accuracy, e.g. pie charts and heat maps (comparing angles and colours).

3.3.1.2 Simplify your visualisation

Data visualisation should convey information in a concise way. Try to tell one story per chart and use small multiples, rather than putting all the data into one chart. For each chart, the number of variable dimensions used in it should not exceed the number of dimensions in the used data.

A chart should make the viewers focus on the important information in data. Unnecessary visual components may distract the viewers. Thus you should simplify your visualisation by removing distractions without affecting the messages that we want to convey.

According to Tufte's theory [4], the distractions ("chart junk") that can be removed include Moiré patterns, grids, chart frames, and meaningless colours. A general rule to check if an element can be removed is that, if the chart can still tell the same story after the element being removed, then it is fine to remove it.

3.3.1.3 Explain your visualisation

Sometimes visualisation cannot tell all the stories in data without the help of texts. Thus it is necessary to use labels and annotations to provide explanations with the viewers. These explanations can be implemented in an on demand style, for example only showing the explanations of a visual component when the viewers interact with it.

3.3.1.4 Clarify uncertainties

Some visualisations (e.g. bar charts) summarise many data values into a few statistical results (e.g. average, median, and mode). Inevitably there is uncertainty in these results, and such uncertainty should be shown in the visualisations. There are different ways to visualise uncertainty [8]. One way is to add necessary information such as sample sizes, standard deviations, standard errors, and confidence intervals into the visualisation. Another way is to visualise the distribution of the data.

3.3.1.5 Start axes with 0

For a chart, it is important to start its y-axis from 0, if possible. Truncated y-axes starting from non-zero values may show distorted presentations of the actual effects in data [7]. If the baseline of a y-axis is not 0, then it should be clarified to the viewers.

3.3.1.6 Use colours carefully

As mentioned above, comparison of colours has the lowest accuracy [2, 5]. Thus colours are not suitable for precise comparison. Additionally, if a chart needs to be exported and printed, the choice of colours should be printer-friendly. The use of colours should also consider colour blind viewers (make sure to test the visualisations using colour blindness simulators), the different interpretations of certain colours in different cultures (e.g. red has negative connotations in the West, and positive connotations in the East), and the viewers' capability of remembering colours.

3.3.1.7 Show differences directly

One common purpose of data analysis is to find the difference between variables. If two variables are presented in a line chart as two curved lines, according to the theory by Cleveland and McGill [2, 5], the comparison between them will be very inaccurate. The reason is that people tend to find the closest distance between two curves rather than to find the vertical distance between them. Therefore, to help the viewers compare variables in line charts (especially for curved lines), an extra chart that directly shows the difference should be provided.

3.3.1.8 Combine interaction into visualisation

Using Interaction in visualisation can help the viewers explore, navigate, and drill down into the subset of data. According to the book by Ward et al. [9], the classes of interaction techniques include:

- Navigation: such as panning, rotating, and zooming
- Selection: such as highlighting, deleting, and modifying
- Filtering: such as eliminating records or dimensions
- Reconfiguring: such as reordering the data
- Encoding: changing graphical attributes such as sizes, shapes, and colours
- Connecting: using connections to link views or objects
- Abstracting/Elaborating: changing the level of details
- Hybrid: combining several of the above techniques

3.3.1.9 Use proximity and similarity to highlight messages

Some encoding components of visualisation may affect the message that they convey [9]. Things need to be reviewed include:

- **Proximity** (as shown in Fig. 16): Objects close to each other are perceived to be in the same group. Therefore proximity should be used carefully and deliberately. If two things do not belong to the same group, do not put them close to each other.

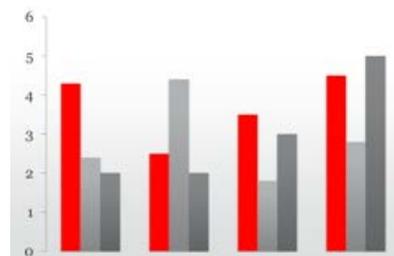


Figure 16. People consider that the bars close to each other belong to the same group, even if they have different colours.

- **Similarity** (as shown in Fig. 17): People also consider that items with the same colour, the same size, or the same shape belong to the same group. Therefore colours, sizes, and shapes should be used carefully and reviewed.

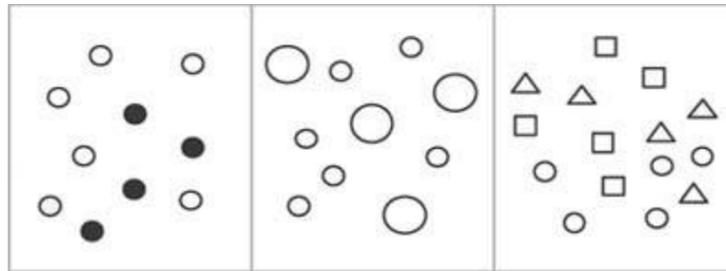


Figure 17. Items with the same colour (left: black and white), the same size (middle: big and small), or the same shape (right: square and triangle) are perceived to belong to the same group.

3.3.1.10 Importance of aesthetics

Apart from being accurate, visualisation should also be aesthetically pleasing. According to the book by Ward et al. [3], the three main aspects to be considered in terms of aesthetics are:

- **Focus:** The viewers' attention should be guided toward the most important part through emphasis and highlight.
- **Balance:** This applies to both dashboards (as discussed in 2.3., through areas with different emphasis) and single charts. For a single chart, the most important information should be put in the centre.
- **Simplicity:** Remove the components that draw attention but do not convey information.

3.3.2 Things NOT to do

3.3.2.1 Using pie charts for comparison

As a technique to present composition, compared with bar charts, pie charts are less accurate as the viewers have to compare angles instead of positions. For example, as shown in Fig. 18, it is difficult to compare the categories close to each other without reading the annotations. Compared with tree maps, pie charts are less capable to present many categories at the same time, as it is difficult to read when the number of categories is higher than 4 (as shown in Fig. 18). Therefore it is not recommended to use pie charts for showing comparison, or showing many parts of a whole. Only use pie charts when showing fewer than 4 parts of a whole, and do not use them for comparison.

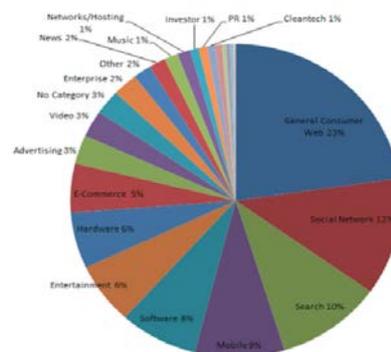


Figure 18. Example of a pie chart with many categories. (Figure from <http://www.storytellingwithdata.com/blog/2011/07/death-to-pie-charts>)

3.3.2.2 3D charts

3D charts (e.g. 3D pie charts, 3D bar charts) present 3D objects on a 2D surface (a screen) through perspective. In perspective, objects close to us are perceived to be bigger than objects far from us, even if their actual sizes are the same. Thus the size of the same visual object may be interpreted differently depending on its position in a 3D chart, which leads to inaccuracy [3]. For example, in Fig. 19 Item C appears to be equivalent to Item A when visualised in a 3D chart, despite being much smaller. Avoid using 3D charts.

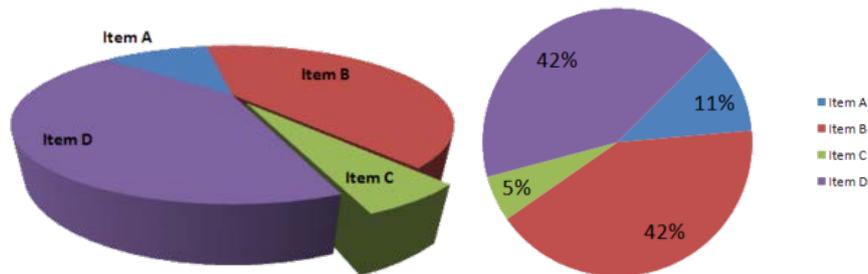


Figure 19. Example of a misleading 3D chart. (Figure from https://en.wikipedia.org/wiki/Misleading_graph)

3.3.2.3 Rainbow colourmap

A rainbow colourmap (red-yellow-green-blue) sometimes is used in heatmaps or choropleth maps to represent different values through different colours. However it is misleading as it introduces artificial colour boundaries even if the actual data is continuous and smooth, which leads to wrong conclusions [3]. For example, in Fig. 20, the transition from 1.0 to 0.5 (red-orange-yellow-green) is actually continuous. But the heatmap shows a “yellow boundary” that clearly divides the red area and the green area. Thus for continuous data, a sequential colourmap should be used. A rainbow colourmap is only safe to use when the data is categorical.

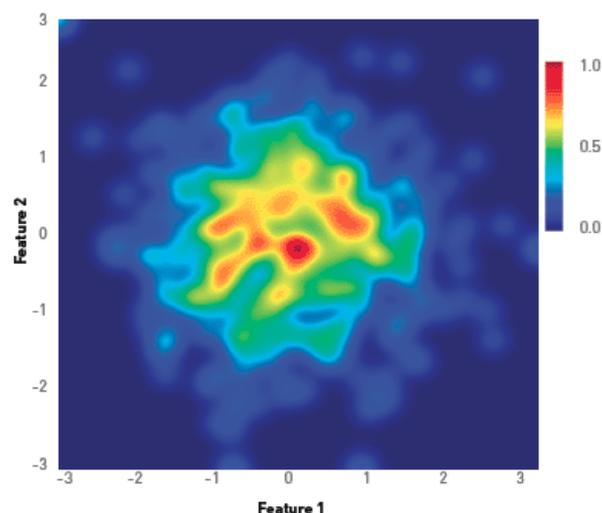


Figure 20. Example of a heatmap using rainbow colourmap. (Figure from [3])

3.3.2.4 Double y-axes

When using double y-axes, the viewers sometimes find it difficult to figure out which data to read against which y-axis. For example, in Fig. 21, the numbers on the two y-axes are similar, which makes

it difficult to find the correct axis for each of the two lines. In addition, the intersection of the two lines is misleading. The viewers may think that it means something meaningful, but in fact it does not make any sense because the two lines do not share the same y-axis. Remember, each visualisation should tell one story (see 4.1.2). Try to avoid using double y-axes and instead consider separating the data into two vertically aligned charts and each of them having its individual y-axis.

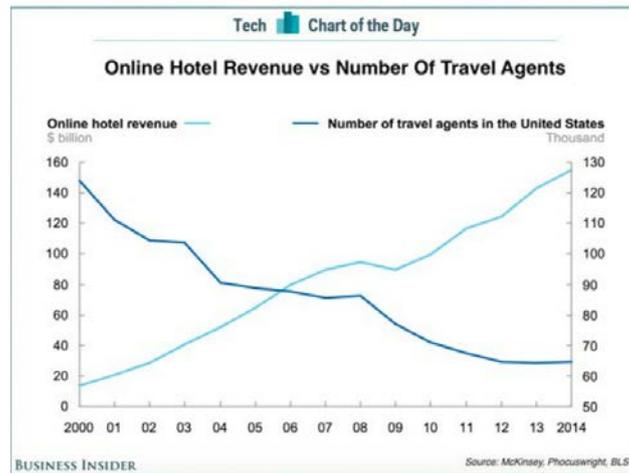


Figure 21. Example of a line chart using double y-axes. (Figure from <https://www.businessinsider.com/travel-agents-vs-online-hotel-bookings-2015-12?r=UK>)

3.3.2.5 Abusing visualisations

Not all information has to be visualised. If the information can be represented using a single number, then you do not need to visualise it. Conversely, in many situations, you should not pack all insights of your analysis into one chart - instead, decide on a series of charts, and arrange them in a coherent sequence.

4 WP4 support for pilots and data stories

In Section 2 and 3, we have discussed the usage of data visualisations in real-world data portals and the theoretical principles of visualisation design. In this section, we discuss how we support the partners for pilots to evaluate whether the visualisations in their business cases are effective. In addition, we summarise our ongoing support for OpenOpps to produce data stories from public procurement data.

4.1 Common tasks of procurement data portal users

One way how WP4 supports the partners is to evaluate the effectiveness of visualisations in their business cases. To understand the tasks that are conducted by the users of the Zaragoza City Council portal, we asked our partner to summarise a number of potential tasks in the format of users' demands or questions. The detailed tasks are as follows:

- Obtain all contracts awarded by the Zaragoza City Council.
- Obtain the contracts that Zaragoza Vivienda has in the bidding phase.

- Obtain the services contracts tendered in 2017 by the Zaragoza Cultural entity
- Obtain the works tendered by the City Council with a bidding amount without VAT of more than 50,000 euros.
- Obtain the contracts in which El Corte Inglés has bid.
- Obtain the contracts in which it has awarded XXX by means of a UTE
- What CPVs does El Corte Inglés usually present itself to?
- Which year was the one with the most contracts tendered by the Networks and Systems Service?
- Which is the contract managed by Ecociudad with more suppliers in 2018?
- Which is the company that has submitted more contracts in 2017? And the one that has awarded more contracts? And the one that has received more money?
- Which companies usually submit contracts together?

4.2 User Studies

Based on the summarised user tasks, we propose a number of potential user studies to investigate how data visualisations can help with these tasks. Each of these studies uses a specific type of visualisations to support a user task. We will evaluate the visualisations between different user groups (i.e., with and without visualisations) and measure the difference between the performances (e.g., response time and accuracy) of the users in different groups. Our proposed studies are as follows.

- Visualising contracts by sectors
 - Question: What CPVs does El Corte Inglés usually present itself to?
 - Generalised Task: Identifying the distribution of data on a certain column
- Visualising contracts over time
 - Question: Which year was the one with the most contracts tendered by the Networks and Systems Service?
 - Generalised Task: Identifying the most frequent value of a certain column
- Visualising contracts by sellers
 - Question: Which is the contract managed by Ecociudad with more suppliers in 2018?
 - Generalised task: Identifying the most frequent value of a certain column
- Visualising tenders by sellers
 - Questions
 - Q1: Which is the company that has submitted more contracts in 2017?
 - Q2: Which is the company that has awarded more contracts in 2017?
 - Q3: Which is the company that has received more money in 2017?
 - Generalised task
 - Q1 & Q2: Identifying the most frequent value of a certain column
 - Q3: Identifying the highest summarised value based on a certain column
- Visualising tenders by co-bidders
 - Question: Which companies usually submit contracts together?
 - Generalised task: Identifying the distribution of the relationships of the values of a column

5 Conclusion

This deliverable has presented the second version of the visualisation and interaction components in WP4. It has discussed how data visualisations are used in the existing public procurement data portals, rules and principles in visualisation design, and how WP4 supports the partners in evaluating data visualisations and making data stories. Audience both within the TBFY consortium and outside would benefit from this deliverable in different aspects.

During the rest of the project, the interface of the TBFY knowledge graph will be developed based on the guideline in this document. In addition, a number of user studies will be conducted according to the proposed plan. The resulting interface and user study results will be delivered in D4.6 in M27.

6 References

- [1] S. Few, *Information Dashboard Design*, O'Reilly Sebastopol, CA, 2006.
- [2] W. S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531-554, 1984.
- [3] D. A. Szafir, "The Good, the Bad, and the Biased: Five Ways Visualizations can Mislead (and How to Fix Them)," *Interactions*, vol. 25, no. 4, pp. 26-33, 2018.
- [4] E. R. Tufte, *The Visual Display of Quantitative Information*, vol. 2, Graphics press Cheshire, CT, 2001.
- [5] W. S. Cleveland and R. McGill, "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data," *Journal of the Royal Statistical Society: Series A (General)*, vol. 150, no. 3, pp. 192-210, 1987.
- [6] D. F. Williamson, R. A. Parker and J. S. Kendrick, "The Box Plot: A Simple Visual Method to Interpret Data," *Annals of Internal Medicine*, vol. 110, no. 11, pp. 916--921, 1989.
- [7] D. Huff, *How to Lie with Statistics*, WW Norton & Company, 1993.
- [8] M. Greis, A. Joshi, K. Singer, A. Schmidt and T. Machulla, "Uncertainty Visualization Influences How Humans Aggregate Discrepant Information," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 2018.
- [9] M. O. Ward, G. Grinstein and D. Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications*, AK Peters/CRC Press, 2015.