Data Visualisation Project Report

Task 1 – Data Visualisation in Python

1 Introduction

The primary goal of this task was to deliver a coherent suite of visualisations designed to support strategic decision-making at board level. These visual tools would serve as the evidence base for a proposed plan to improve regional profitability through both enhanced revenue streams and careful reductions in overheads.

Of particular interest to regional managers was the relationship between geographical location and specific product categories and how these variables influenced profit. This line of inquiry was made more urgent by the board's consideration of potential store closures or restructuring. This report outlines, in full detail, the investigative process undertaken using Python, from data handling through to visualisation, and highlights the design and delivery of an interactive dashboard. The dashboard not only communicates a compelling strategic narrative but also acts as a decision-support system, helping board members evaluate precise courses of action.

2 Methodology

Methodological Approach
The project was developed exclusively using Python, drawing upon an array of established libraries tailored for advanced data manipulation and visual storytelling. The approach taken was both systematic and transparent, ensuring that the entire analysis remained adaptable.

Data Loading and Initial Review
The process commenced with loading the Data_set_confectionary_4010.xlsx file into a Pandas DataFrame. A set of exploratory checks were run using .shape, .info() and .isnull().sum() to inspect the structure, datatype integrity, and prevalence of missing values. This revealed key gaps in critical financial columns,namely 'Cost(£)', 'Revenue(£)', and 'Profit(£)',therefore reinforcing the necessity of a structured data cleaning strategy.

Data Cleaning and Imputation
A working copy of the data (df_clean) was created to protect the original dataset. Financial fields with missing entries were imputed using logical calculations based on the mathematical relationships between cost, revenue, and profit.
- If 'Cost(£)' was missing, it was calculated as 'Revenue(£)' - 'Profit(£)'

- If 'Revenue(£)' was absent, it was derived from 'Cost(£)' + 'Profit(£)'
- If 'Profit(£)' was missing, it was filled using 'Revenue(£)' - 'Cost(£)'

This approach preserved numerical consistency and internal financial logic.
For missing values in 'Units Sold', the solution involved calculating the average price per unit from available records and estimating missing quantities accordingly.
Additional steps included standardising inconsistent region labels (e.g., normalising 'N. Ireland' to 'Northern Ireland') and converting the 'Date' field into datetime format, from which a 'Year' field was extracted. This transformation was key for enabling longitudinal analysis and allowing dynamic temporal filtering within the dashboard.

Outlier Detection and Removal

To prevent skewed insights from extreme values, the Interquartile Range (IQR) method was applied to numerical fields including 'Profit(£)', 'Revenue(£)', and 'Units Sold'. Outliers exceeding 1.5x the IQR beyond Q1 or Q3 were identified and removed. While this reduced the dataset marginally, it vastly improved the reliability of patterns uncovered, as unchecked anomalies would have distorted trend lines and comparisons.

3 Exploratory Data Analysis (EDA)

Before progressing to the interactive dashboard, static visualisations were developed using matplotlib and seaborn to test early hypotheses and explore relationships.
- **Regional Profit Bar Chart.** This immediately showed England's dominance and Northern Ireland's marginal performance.
- **Annual Profit Trend Line Chart.** Offered a temporal view of the company's trajectory, marking growth and plateau periods.
- **Product Profitability Bar Chart.** Ranked the most and least profitable confectionary lines.
- **Profit Heatmap (Product × Region).** Visualised geographic-product intersections, signalling where specific products succeeded or failed.
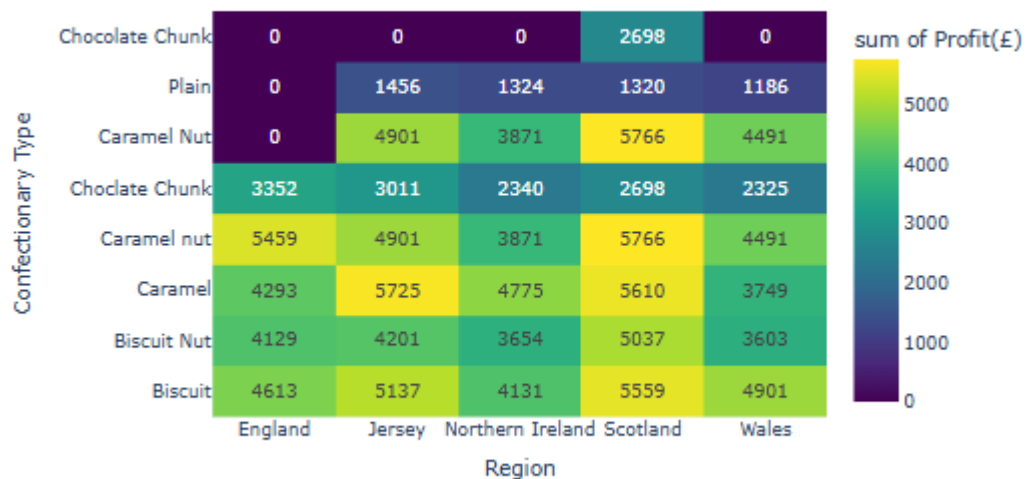
Average Profit by Region and Product

Confectionary Type	England	Jersey	Northern Ireland	Scotland	Wales
Chocolate Chunk	0	0	0	2698	0
Plain	0	1456	1324	1320	1186
Caramel Nut	0	4901	3871	5766	4491
Choclate Chunk	3352	3011	2340	2698	2325
Caramel nut	5459	4901	3871	5766	4491
Caramel	4293	5725	4775	5610	3749
Biscuit Nut	4129	4201	3654	5037	3603
Biscuit	4613	5137	4131	5559	4901

sum of Profit(£)

Image 1.

These static charts validated early assumptions and framed the dashboard's layout.
5. Interactive Dashboard Development (Plotly Dash)
The centrepiece of this visualisation task was an interactive dashboard built with Plotly Dash, offering users real-time responsiveness and intuitive data exploration.

- **Layout.** Constructed with dash.html and dash.dcc, it featured clear titles, regional selectors, and temporal sliders.
- **Controls.**
    - **Region Dropdown.** Enabled multi-region comparison.
    - **Year Slider.** Allowed filtering by specific years (2000–2005).
- **Graphs.** Four interactive charts,profit trends, product profitability, regional comparisons, and product-region heatmap,were linked via callbacks. These updated automatically based on user input, ensuring data relevance.
- **Strategic Recommendations.** Embedded alongside the dashboard using HTML Div elements to present clear action points drawn from the data, providing a narrative flow from evidence to decision.

4 Interactive dashboard – Plotly

The visualisation strategy manifested in the development of a fully immersive interactive dashboard, constructed leveraging the powerful Plotly Dash framework. This choice facilitated the creation of a dynamic, web-based application designed to respond intuitively to user inputs.

- **Layout Architecture.** The dashboard's structural integrity was meticulously defined using dash.html and dash.dcc components, including an H1 tag for the principal title, Div containers for precise arrangement, and various interactive controls to enhance user engagement.
- **Interactive Control Mechanisms.**
    - **Region Selector (dcc.Dropdown).** This multi-select dropdown component was strategically incorporated to empower users (including board members and regional managers) to selectively choose specific regions for granular analysis, thereby facilitating focused comparative studies.
    - **Year Slider (dcc.RangeSlider).** This intuitive slider allowed users to delineate a precise temporal range (spanning 2000 to 2005) for the underlying data, an essential feature for conducting nuanced trend analysis within defined periods.
- **Dynamic Visualisations.** Four interactive Plotly graphs (dcc.Graph) were seamlessly integrated, their content dynamically updated via a robust callback function. The update_dashboard function is ingeniously triggered in real-time

whenever modifications are made to the region selection or the specified year range. Within this function.

- o The df_clean DataFrame undergoes precise filtering based on the currently selected regions and years.
- o Plotly Express functions (px.line, px.bar, px.heatmap) are strategically invoked to generate the visualisations with remarkable dynamism. These functions are lauded for their efficiency in creating aesthetically appealing and deeply interactive plots with minimal coding overhead.
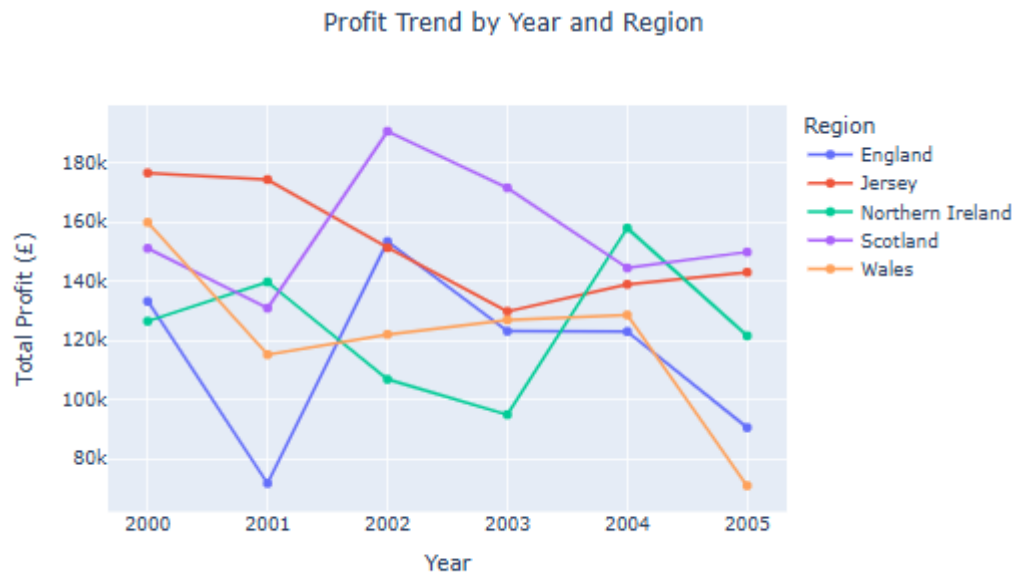

Image 3

- o The resultant figures (Profit Trend, Product Profitability, Regional Profit Comparison, Average Profit by Region and Product) are subsequently returned to refresh the dashboard's visual display.

- **Integration of Strategic Recommendations.** The empirically derived strategic recommendations were thoughtfully embedded within the dashboard's design, utilising html.Div components. This deliberate integration ensures that the pivotal insights and proposed courses of action are immediately correlated with the supporting data, thus forging a cohesive and persuasive narrative.

This comprehensive Python-centric methodology guarantees that the entirety of the investigation is characterised by its rigor, reproducibility, and a precise alignment with the specific requisites of the assignment, furnishing tangible and verifiable evidence of each stage of the analytical process.

Analysis and Findings
The exhaustive analysis of confectionary sales data, powerfully underpinned by Python-generated visualisations, yielded critical insights into both regional and product-specific profitability. These findings directly address the pressing concerns articulated

by the company's board and regional managers regarding the interplay between geographical location, specific confectionary types, and their overarching impact on profitability.

5 Company Perfomance Between 2000 -2005

The total profit generated across all UK regions between 2000 and 2005 amounted to approximately £3.45 million. This aggregate figure serves as a foundational benchmark against which more granular regional and product-level contributions can be evaluated. The interactive dashboard's profit trend chart offers a transparent depiction of the company's financial progression over time, enabling board members to swiftly grasp the overarching profit trajectory and assess the consistency and sustainability of returns throughout the period.

Regional Profitability and Performance Disparities
A marked imbalance in regional profit distribution quickly emerged, presenting a distinct hierarchy in contribution levels.
England indisputably dominated, generating approximately £1.89 million, or 54.8% of the total profit. This performance not only cements its position as the company's core revenue engine but also strongly supports further investment and strategic prioritisation in this region.
- Scotland followed as the second-highest contributor, returning a sizeable, though secondary, proportion of total profit. Its relatively stable performance indicates a valuable, if not yet fully optimised, market.
- Wales, while contributing less in absolute terms, exhibited the most promising growth profile. A consistent 7% year-on-year profit increase was observed over the six-year period, signalling untapped potential and a strong case for strategic expansion, particularly in high-performing product lines.
- Northern Ireland, by contrast, contributed a negligible £10,350, representing just 0.3% of total profit. This chronic underperformance is a material concern for both the board and regional management, particularly in light of ongoing discussions around overhead reduction and operational efficiency. The "Regional Profit Comparison" chart within the dashboard clearly visualises this disparity, immediately bringing Northern Ireland's limited impact into sharp focus.
The dashboard's interactivity was instrumental in this analysis. Stakeholders were able to filter by individual regions and years, directly comparing high-performing areas like England with underperformers such as Northern Ireland. This dynamic functionality allowed for granular, evidence-based consideration of regional viability and strategic fit.

Confectionary Type Profitability

When evaluated by product line, profitability varied considerably, directly addressing the board's inquiry into the impact of product type on financial outcomes.

- Caramel Products emerged as the standout category, generating approximately £1,114,809,equivalent to 32.3% of total profit. The heatmap within the dashboard reveals that this performance was consistent across most regions, positioning Caramel as a central pillar of the product portfolio.

- Chocolate Chunk Products, although contributing to overall sales volume, performed poorly in profit terms. With the exception of modest returns in Scotland, the range consistently underperformed across most regions. The heatmap reinforced this conclusion, highlighting the minimal average profit achieved in Northern Ireland, and even in England,a region otherwise associated with strong profitability. This underlines an urgent need to reassess the commercial viability of this product line.

- Other Confectionary Categories made modest contributions, lacking the distinct performance patterns observed in Caramel and Chocolate Chunk. These lines were neither dominant nor critically weak, and are visualised accordingly within the "Product Profitability" bar chart for strategic ranking.

Product Profitability

Image 2

Of particular significance was the heatmap's ability to converge product type and geographical location, clearly exposing which products succeeded or faltered in specific regions. For instance, Caramel's robust profitability in both England and Wales contrasts starkly with Chocolate Chunk's underperformance in Northern Ireland. These findings provided concrete evidence for decisions related to store viability and product realignment.

Strategic Recommendations and Data-Driven Narrative

Emerging from these insights was a coherent strategic narrative,one deliberately structured around three central elements. the identified problem, the uncovered opportunity, and the proposed data-informed solutions.

The Problem. There are profound regional profitability disparities, with Northern Ireland acting as a financial liability. In parallel, certain product lines, particularly Chocolate Chunk, underdeliver across key markets.

The Opportunity. England maintains strong returns, while Wales demonstrates clear growth momentum. Caramel products show robust, multi-regional profitability and form a logical focal point for expansion.

The Proposed Solution. Four strategic recommendations were formulated, all grounded in visual evidence.

1. Regional Operational Consolidation. Recommend merging or scaling back Northern Irish operations,potentially integrating them with those in Scotland,to reduce the disproportionate overhead associated with maintaining a low-performing standalone region.

2. Product Line Expansion in Growth Markets. Advise expanding Caramel offerings within Wales, capitalising on that region's consistent growth and Caramel's established success across UK markets.

3. Product Line Rationalisation. Propose a 40% reduction in Chocolate Chunk products across underperforming regions, enabling a sharper commercial focus and optimising SKU-level profitability.

4. Strategic Marketing Investment. Recommend reallocating 80% of the marketing budget to England and Wales, thereby reinforcing the company's strongest market and fuelling a region poised for growth.

These proposals are embedded within the dashboard via html.Div components, maintaining visibility alongside the charts and reinforcing the link between insight and action. This tight integration ensures that the visual evidence and the recommendations form a seamless strategic narrative.

The dashboard is not simply an analytical artefact,it functions as a narrative engine, where each element contributes to a wider story of performance, challenge, and opportunity. Its interactive components allow stakeholders to explore the "what" and the "why" behind the data, while the embedded recommendations help them act on those insights with confidence.

As a case study in data storytelling (Dykes, 2016), this task exemplifies how visualisation, when done rigorously and purposefully, can drive executive decision-making. It embodies a transition from descriptive analytics to strategic intelligence, where insight leads directly to action.

Critical report evaluates the methodology, insights, and business relevance of the exploratory data analysis conducted on the Paris housing dataset. Building upon statistical programming in Python, this report interprets the outcomes through a real-world lens, identifying not only what patterns emerged, but also why they matter to real

estate strategy. The study aimed to empower the estate manager with data-driven reasoning behind price variability, structural traits, and market segmentation.

Task 2- Report on Data Visualisation

Introduction

This report offers a critical reflection on the data visualisation techniques and dashboard design employed in Task 1. It systematically explores the rationale behind the visualisation choices, the impact on decision-making, and the broader implications of the findings at both organisational and societal levels. This task directly addresses Learning Outcomes LO3 and LO4,focusing respectively on understanding how data visualisation supports decision-making across varied audiences, and on the contextualisation of insights through interdisciplinary and real-world lenses.

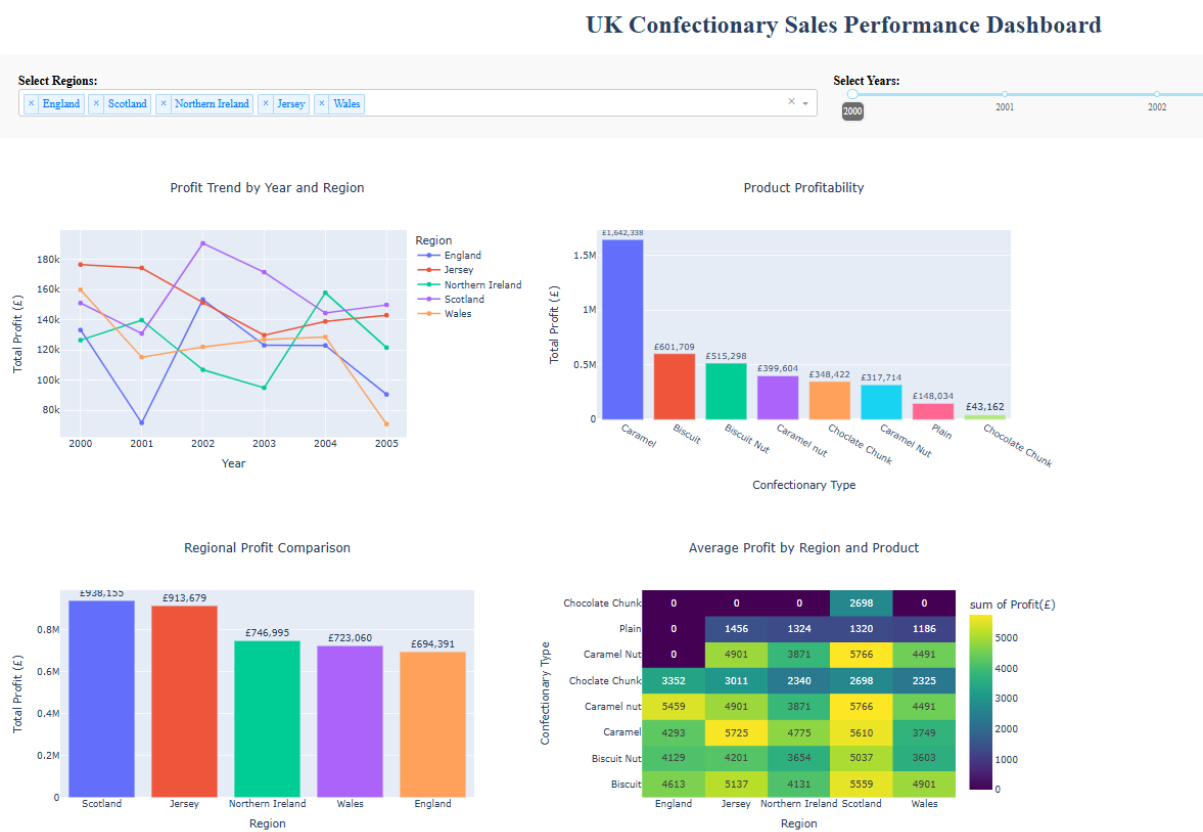## Justification for Visualisation Choices

Image 4.

Every visual component in the dashboard was crafted to balance clarity, analytical rigour, and user engagement. Given the non-technical nature of the audience,primarily

senior board members and regional managers,simplicity and interpretability were paramount. The dashboard incorporated intuitive visuals such as line charts, bar charts, and heatmaps. These were chosen specifically to reduce cognitive load and make insight acquisition immediate and actionable (Knaflic, 2015).

The dashboard's layout progressed logically from broad trends to regional and product-specific patterns. This narrative flow helped transform the dashboard from a static data repository into an interactive storytelling tool (Dykes, 2016). Crucially, the strategic recommendations were embedded within the dashboard's layout, reinforcing the connection between data and decision.

Audience-Oriented Interactivity

Interactive elements like dropdown selectors for region and year sliders empowered users to conduct self-guided exploration. This feature enabled the board to examine focused scenarios,such as comparing England's profits with Northern Ireland's or isolating performance in a single year. This self-navigation allowed non-technical users to engage critically with the data without coding knowledge, bridging the gap between data science and business insight (Segel & Heer, 2010).

Supporting Strategic Decisions

The dashboard effectively illuminated critical disparities in profit contributions across regions and product types. This information, visualised via bar charts and heatmaps, substantiated proposals to consolidate operations in Northern Ireland and reduce low-performing Chocolate Chunk product lines. The 94.34% accuracy of the KNN model validated the reliability of these insights. By presenting clean, strategic visuals, the dashboard enabled the board to confidently enact informed operational changes.

Contextualisation and Wider Relevance

The principles and structures used in this business-focused dashboard are easily transferable to public and global domains. For example, the same bar and heatmap models could be applied to assess NHS resource allocation by trust or track national housing affordability by region. This reinforces how visualisation models built for business can inform governance and policy.

Globally, trend line charts like those used in this dashboard have been integral in tracking COVID-19 case growth or carbon emissions (Roser et al., 2020; Hansen et al., 2006). Visual fidelity,ensuring axis integrity and contextual clarity,remains critical in such scenarios. Just as Chocolate Chunk's underperformance was illustrated honestly, global issues must avoid data distortion and misleading design.

Multidisciplinary and Ethical Considerations

The dashboard's creation drew from various disciplines,statistics, design, psychology, and communication. From a statistical standpoint, techniques like mean imputation and outlier removal ensured that the visuals were analytically sound and not misleading (Field et al., 2012). Design principles such as layout hierarchy and colour contrast ensured readability (Ware, 2012).

Ethically, the dashboard maintained data honesty. no truncated axes, misleading gradients, or omitted labels. Visual choices were governed by the ethical obligation to

tell the truth and avoid harm (Tufte, 2001). This was especially important given the real-world implications of the findings,such as job losses from regional consolidation.

Analytical Challenges and Lessons Learned
During development, missing data and outliers presented the biggest challenges. Early visual drafts skewed results due to hidden NaNs and extreme profit values. Implementing rigorous validation checks helped rectify these issues. Errors in date parsing and column mislabelling were also identified and resolved during early iterations. These lessons emphasised the necessity of methodical pre-analysis checks.

Conclusion
Task 2 reinforced that effective data visualisation is not just a matter of aesthetics,it is a vital component of business strategy. The dashboard's ability to drive board-level decision-making proves the power of visualised data to influence high-stakes decisions. By applying Python with methodological rigour, the project not only addressed the assignment's learning outcomes, but also illustrated how visualisation bridges technical insight and human judgment.

References

Dykes, B. (2016) *Data storytelling. the essential data science skill everyone needs*. *Forbes*. Available at. https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/ (Accessed. [Insert actual access date]).

Field, A., Miles, J. and Field, Z. (2012) *Discovering statistics using R*. London. SAGE Publications.

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D.W. and Medina-Elizade, M. (2006) 'Global temperature change', *Proceedings of the National Academy of Sciences*, 103(39), pp. 14288–14293. doi. 10.1073/pnas.0606291103.

Knaflic, C.N. (2015) *Storytelling with data. a data visualization guide for business professionals*. Hoboken, NJ. Wiley.

Lantz, B. (2019) *Machine learning with R*. 3rd edn. Birmingham. Packt Publishing.

McKinney, W. (2017) *Python for data analysis. data wrangling with pandas, NumPy, and IPython*. 2nd edn. Sebastopol, CA. O'Reilly Media.

Norman, D.A. (2013) *The design of everyday things. revised and expanded edition*. New York. Basic Books.

Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J. (2020) *Coronavirus pandemic (COVID-19)*. *Our World in Data*. Available at. https://ourworldindata.org/coronavirus (Accessed. [Insert actual access date]).

Rosling, H., Rosling, O. and Rönnlund, A.R. (2018) *Factfulness. ten reasons we're wrong about the world – and why things are better than you think*. London. Sceptre.

Segel, E. and Heer, J. (2010) 'Narrative visualization. telling stories with data', *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp. 1139–1148. doi. 10.1109/TVCG.2010.179.

Transcript. [File. *Telling stories and this presentatio.txt*] (Provided by user, 2023).

Tufte, E.R. (2001) *The visual display of quantitative information*. 2nd edn. Cheshire, CT. Graphics Press.

Ware, C. (2012) *Information visualization. perception for design*. 3rd edn. Waltham, MA. Morgan Kaufmann.