INFS 5116 – Data Visualisation

# Practical Exercise 2 – Create your own graphics using R

**Student name**:

**Student ID number**:

**Introduction**

Provide a brief overview of the chosen supermarket product(s). Why did you choose it? What are you trying to find out?

Also include any other relevant background information regarding data preparation and cleaning.

One short paragraph will be sufficient.

**Highlight and delete this message before submission**

[Type your introduction here]

**Graphic 1**

Use the table below to present your graphic, describe your design choices and propose some questions and conclusions.

Be brief; you can use dot points if you wish. Repeat for the rest of your graphics.

**Highlight and delete this message before submission**

|  |  |  |
| --- | --- | --- |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | Plot 1  Level of Analysis: Elementary  Multi-Component: No  Design Principles: The design employs a bar chart where the x-axis is the 'Store' and the y-axis is 'Frequency'. The color scale is viridis, which provides a perceptually uniform scale. theme\_ipsum() and theme\_minimal() are applied for aesthetic simplification.  Questions Answerable: What is the frequency distribution of each store? Are there stores that are particularly over or underrepresented in the dataset? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | Second Plot: Top and Bottom 10 Stores by Sales  Level  The graph falls under the "**Intermediate-Level Question Visualizations**" category. It does not capture the entire database but provides significant insights into specific stores' performance based on total unit sales.  Components  It's a multi-component plot as it compares two categories: the top 10 and bottom 10 stores.  Design Principles  Color Mapping: Price\_Tier is used as the fill color to distinguish different pricing categories.  Scale: Y-axis is scaled to the actual sales numbers, providing a sense of volume.  Layout: Utilizes side-by-side bar graphs for direct comparison.  Questions Answered  Which stores are performing the best and worst in terms of unit sales?  How does the Price Tier correlate with total sales for each store? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |

|  |  |  |
| --- | --- | --- |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | 100% Stacked Histogram of Gross Margin by Price Tier  The third plot belongs to **the Intermediate-Level** Question Visualizations category. The graph doesn't encapsulate the entire database but conveys significant information about how profit varies by Price Tier.  Components  The graph is a single-component visualization. It portrays a binned distribution of the gross margin (PROFIT) segregated by different Price\_Tiers.  Design Principles  Data Encoding: The x-axis encodes the profit bins (binned\_profit), and the y-axis encodes the proportion (prop) of each bin per Price Tier. The fill color represents the different Price Tiers.  Data Transformation: The gross margin (PROFIT) has been binned into discrete ranges for easier interpretation and visualization.  Data Integrity: Removed rows with NA values to maintain data integrity.  Color Scheme: The Viridis color palette was used to differentiate between Price Tiers, which is effective for those with color vision deficiencies.  Positioning: The stacked histogram uses vertical bars to illustrate the distribution of PROFIT within each Price Tier, giving a sense of proportion.  Example Questions  What is the proportion of high-profit products within each Price Tier?  Which Price Tier has a higher percentage of low-profit or negative-profit items?  Is the distribution of profit similar across different Price Tiers? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |

|  |  |  |
| --- | --- | --- |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | Heatmap of Profit by Month  Classification  The fourth plot can be categorized under **Overall Question** Visualizations, particularly in the Heatmap subcategory. It offers an overview of how gross margins (PROFIT) are distributed across months (monthf).  Components  This is a single-component visualization. It displays the relationship between two numerical variables: monthf and PROFIT, while color encodes a third metric, density (..density..).  Design Principles  Data Encoding: The x-axis represents the month (monthf), and the y-axis signifies the gross margin (PROFIT). The color density indicates the frequency of data points.  Data Transformation: geom\_bin2d() is used to create bins that classify the data points. This binning approach simplifies the data structure and allows for visual examination of densities.  Color Scheme: A Viridis color palette was applied to indicate density. The colors help identify regions where data points are concentrated.  Aesthetics and Theme: A minimalistic theme is applied, which helps focus on the data rather than the embellishments.  Positioning: Data points are binned in a 2D space based on both PROFIT and monthf. This format helps to immediately identify the concentration of higher or lower profits across different months.  Example Questions  Which months have the highest density of higher-profit items?  Is there any month where low or negative-profit items are more frequent?  Are there any significant patterns that can be observed when profits are mapped against months? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |

RDI plots or treemaps

|  |  |  |
| --- | --- | --- |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | Gross Margin by Price Tier and Year  Classification  The treemap falls under the category of Intermediate-Level Question Visualizations. This plot aims to demonstrate hierarchical relationships and part-to-whole relationships, focusing on how profits vary across different years and Price Tiers within the dataset.  Components  This is a single-component visualization showing the structure of the 'PROFIT' metric segmented by 'Price\_Tier' and 'year'.  Design Principles  Data Encoding: The rectangles' area represents the 'PROFIT' while the color encodes the 'Price\_Tier'. The nested structure provides a hierarchical view.  Data Transformation: Grouping is performed by 'Price\_Tier' and 'year', with the latter nested within the former. Summation of 'PROFIT' within these groups forms the basis for the rectangles' size.  Data Integrity: There are no indications of missing or outlier data affecting the integrity of this treemap.  Color Scheme: The "Blues" color palette was used to differentiate between Price Tiers. This palette is perceptually uniform and provides good contrast.  Positioning: The treemap employs spatial positioning and enclosure to represent hierarchy and size, making it easy to compare categories visually.  Example Questions  Which 'Price\_Tier' consistently has higher 'PROFIT' across years?  Is there a year where a particular 'Price\_Tier' experienced a significant spike or drop in 'PROFIT'?  How does 'PROFIT' distribution within a 'Price\_Tier' change over years? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |

|  |  |  |
| --- | --- | --- |
| **Design information** | **Graphic** | **Questions of interest and conclusion** |
| Type of graphic (e.g. bar chart)  Invariant  Components (number and type)  Imposition  Planar variables  Retinal variables  Gestalt principles (if relevant)  Bertin’s schemata (optional) | Calendar Heat Map: Retail Price by Weekday, Month, and Year  Classification  The plot falls under the category of Time-Series Heatmaps and addresses overall questions. It aims to visualize the variation in retail price across weekdays, months, and years.  Components  This is a single-component visualization that displays the metric 'PRICE' as the main focus. The variables 'year', 'month', and 'weekday' serve as axes to segment the data.  Design Principles  Data Encoding: The color intensity represents the 'PRICE', with higher prices corresponding to a darker shade. The grid is divided based on 'year' and 'month', providing a time dimension.  Data Transformation: The data is grouped by 'year', 'month', and 'weekday'. Within each group, the 'PRICE' is represented as a color gradient.  Data Integrity: Assuming the dataset is complete and accurate, there are no indications of missing or outlier data that would compromise the visualization's integrity.  Color Scheme: A gradient from light gray ("#939597") to red ("#c80000") is used to differentiate price levels. This choice is visually impactful but might lack the perceptual uniformity of other palettes.  Positioning: The calendar heatmap uses spatial positioning to differentiate between time units, making it easy to compare price trends over different periods.  Example Questions  How does retail 'PRICE' vary within a week for each month and year?  Are there any patterns or anomalies in 'PRICE' across weekdays?  Is there a noticeable seasonal variation in 'PRICE'? | [You do not have to cover all reading levels with every graphic, aim for questions most relevant to your graphic.]  Questions:  Conclusion: |