COMP0034 - CW1

Visualization Design

Personas

Persona of general public from COMP0034 CW1:

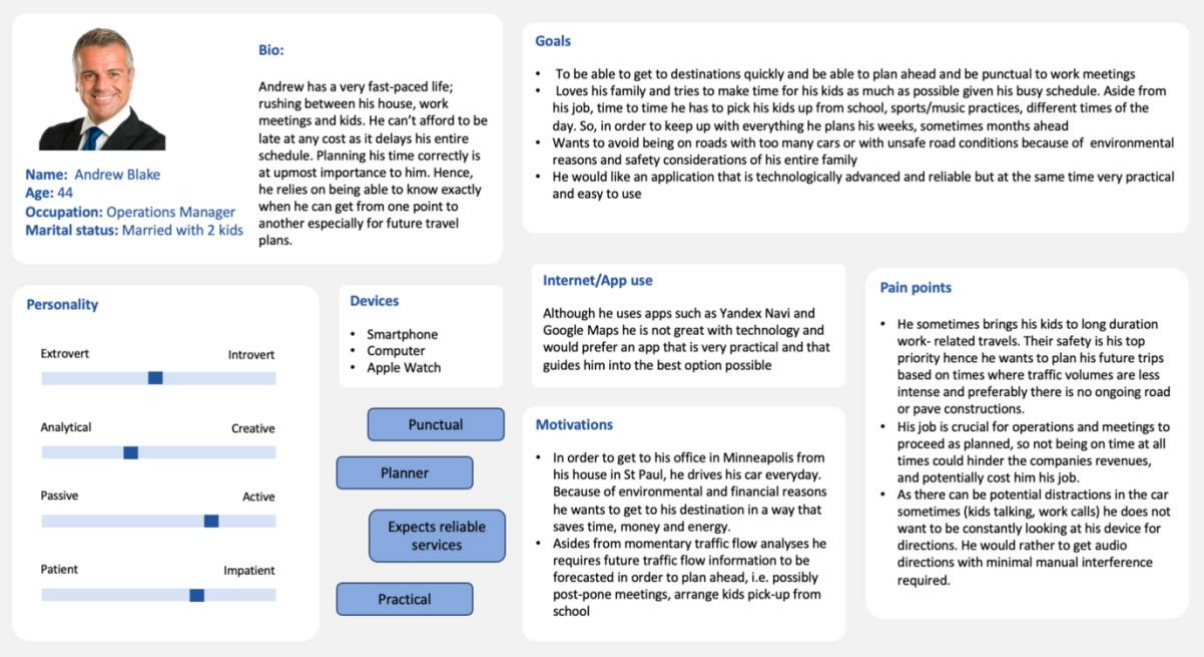

Figure 1:User Persona of General Public

New persona of 'environmentalist policy-maker / researcher:

Note: while creating the app I have decided to dedicate a page to especially researchers, so I thought there was great value to create another persona which would likely benefit from app 2's functionalities

Figure 2:User Persona of Environmental Policy-Maker / Researcher

Data science questions

- What is the distribution of traffic volume over the years? How did traffic volume change over time; are there any seasonal, annual patterns or trends?
- Do certain features show consistent and periodic fluctuations?
- Which features have the most impact on traffic volume (most reliable predictors)?
- Which hours of the day yield the highest traffic volume?
- What is the effect of special days (holidays) on traffic volume?
- What is the effect of weather features on traffic volume? Which weather yields the highest traffic volume?
- Are there any outliers that skew the distributions? Does the data have complete records?
- Are some of the variations observed simply random changes?
- Are the correlations between attributes intuitive?

Within assets folder of my repository, I have included the Business Need file (from COMP0034 CW1) and Requirements file (from COMP0034 CW2) to allow markers assess whether the design is appropriate for the audience.

Remark: some of the designed functionalities of the app in these files significantly changed this term

Dashboard and Visualization Design

Selection of Chart Types

While designing the dashboard I faced a challenge as I didn't want to hide any relevant data, but at the same time considering the principle of 'less is more effective &attractive' from Dark Horse Analytics (Cherdarchuk, 2013), I didn't want to present too much data, (as in lots of noise, salient data would be challenging to identify). Hence, I tried to create an interactivity and dynamically updating platform in which users can manipulate/ the content they see. To give users more control and ability to curate the content they see, I used several dropdowns, radio items and checklists. The interactivity provokes user engagement and allows users to discover patterns and relationships between multiple features but also allows users to choose to ignore features which they believe are less important or relevant. Considering the target audience of both general public and environmental policy makers and their needs from the app, I aimed to make the main view of the dashboard and the visualizations more familiar, easier to understand and prioritize giving the users control over what is displayed, switch charts, access different elements and observe relationships.

I wanted to choose graphic types that matches and supports the nature of my data, the message I want to convey to my target audience and purpose

app1.py

To show distribution and tendencies over time, in the **1st tab** of app 1, I choose to present an option of 2 distribution plots; a box plot and a violin plot (traffic volume over date time features).

A violin plot is a hybrid of a box plot and a kernel density plot. Just like a box plot, it is used to visualize the distribution of numerical data and shows peaks in the data. However, unlike a box plot which can only show summary statistics, violin plots also depict the probability density of each variable' (distribution of the sample data (density trace)). A Box Plot is a common way of visually displaying the data distribution through their quartiles. Though a box plot may seem primitive in comparison to violin plots, they are more straightforward to interpret and take up less space (useful when comparing distributions between many groups)

"Box Plots are limited in their display of the data, as their visual simplicity tends to hide significant details about how values in the data are distributed. For example, with Box Plots, you can't see if the distribution is bimodal or multimodal. While Violin Plots display more information, they can be visually noisier than Box Plots." (The Data Visualisation Catalogue-Violin plot, no date)

There are trade-offs between preference of box and violin plots, meaning, whether if a box plot or a violin plot is more straightforward is subjective. Hence, assuming that while researcher would benefit from density information's provided in violin plots, general public may prefer the simplicity of box plots, I decided to present both options of distribution plots.

To display quantitative values over a time period, analyse how traffic volumes changed over time and reveal trends (annual & seasonal), in the **2nd tab** of app 1, a line chart is generated. In this tab, date time features such as year month day and hour are aggregated to visualize a smoother/clearer pattern, this allows capturing trends and seasonality patterns of traffic volumes.

app2.py

While public may tend to use a mobile phone to get quick information's, researchers and policy-makers would likely use a desktop for more thorough analysis, making more use of the 'add chart' functionality. Considering that the types of devices the target audience is using the dashboard application vary, I wanted to organize the layout of the app such that it suits and is easy to use with both a mobile phone layout and a computer layout. To support the needs of researchers, I wanted to generate multiple different charts but also wanted to figure out a way that avoids fitting all of them in a single view of the dashboard. To solve this issue, in app 2 I created a button for researchers to add as many charts as the they want from their computer. But as the button only adds additional charts if the user prompts, it also supports mobile phones (dynamically updated)

app 2 provides 3 options for type of charts: bar chart, line chart, pie chart. (As each chart can its downsides, I wanted to provide more alternatives). These charts are intended for comparison of impacts of different categorical/ descriptive features on traffic volumes. While a bar chart uses bars to display discrete numerical comparisons across categories, pie chart shows proportions and percentages between categories. Though pie charts are ideal for giving the users a quick overview of the proportional distribution of the data, they are limited show only few values (because as the number of values shown increases, the size of each segment becomes smaller becoming visually unappealing). In app 1, an aggregate line graph is plotted over different date features, while a single line was allowed in app 1, line chart in app 2 allows lines to be grouped with other lines (other data series). So besides revealing trends over time, within a line chart individual lines can be compared to one another

Dynamical callbacks allow linking of these graphs and dropdown options to the number of clicks on the 'add chart' button. (Plotly-Pattern-matching callbacks, no date) (More detailed explanation of how I used 'MATCH' to incorporate this interactivity is provided in README.md of my repository)

Selection of Chart Design & Stylings

In my dashboard application I tried to adhere to the visualization design principles from Alberto Cairo, Edward Tufte, and several others. Following Edward Tufte's views about data-ink ratio & advice on minimizing chartjunk and noise (Tufte, 2001) I avoided non-data ink with unnecessary elements that are not essential to convey the main messages (chartjunk).

"Chartjunk promoters imagine that numbers and details are boring, dull, and tedious, requiring ornament to enliven. Cosmetic decoration, which frequently distorts the data, will never salvage an underlying lack of content. The operating moral premise of information design should be that our readers are alert and caring; they may be busy, eager to get on with it, but they are not stupid."
(Cairo, 2012)

To reduce the noise in my visualizations, I removed axis lines, grid lines, legends, unnecessary labels, and titles. I have also minimised the use of colour to only when it has a particular meaning (colour is used only when grouping elements) and set background colours for all my visualizations to transparent to highlight only essential information.

Following Alberto Cairo's principles on visual lies (Cairo, 2014), to avoid misleading data visualizations and distorting the data I made sure that graphic forms were not used in appropriate ways. For example, all bar plots start with y-axis at zero so that certain data/differences won't be emphasized. I also avoided unnecessary bolding or styling of elements that suggests a particular significance.

Overall, in my dashboard design, instead of more complex and noisy charts which may mislead or could be misinterpreted, to extract meaning from it as easy as possible, I chose relatively more straightforward and intuitive charts. I prioritized enriching the visualisations with interactivity so that users can personalize and organize the content. I tried to give as much control of the information as possible so that users can add level of complexities (more features) only at their will – this generated a platform that is suitable for research purposes as well as personal purposes.

Considering Alberto Cairo's views on trade-offs in visualization, I assumed that my target audiences are likely to prefer different types of visuals. (Cairo, 2012) In this regard, I tried to give the users as much control as possible over the content they see. Through interactivity, users can choose if they I wish to visualise more dense visuals by adding more features, or lightweight visuals by isolating features.

originality vs. **familiarity:**

Trying to address the needs of general public, I tried to create familiar visuals that are easy to understand (message is guaranteed to be conveyed straightforwardly). All the charts provided are more commonplace, familiar visuals for all users to easily grasp and comprehend the information provided and by increasing dynamical content, incentive them to interact/play with different features.

multi-dimensionality vs. **uni-dimensionality:**

considering the target audience of researchers, I tried to create multi-dimensional visuals that are more dense and provide multiple functionalities (specifically app 2). As multi-dimensional visuals illustrate many different aspects of a phenomenon (impact of different features on traffic volume) it gives users the opportunity to dig into the data quite deeply and see the same data in different ways

References:

Cairo, A., 2014. Graphics lies, misleading visuals: Reflections on the challenges and pitfalls of evidence-driven visual communication. In *New challenges for data design* (pp. 103-116). London: Springer London. Available at: https://faculty.ucmerced.edu/jvevea/classes/Spark/readings/Cairo2015_Chapter_GraphicsLiesMisleadingVisuals.pdf (Accessed: February 8, 2023).

Cairo, A., 2012. *The Functional Art: An introduction to information graphics and visualization*. New Riders. Available at: https://www.oreilly.com/library/view/the-functional-art/9780133041187/ch03.html#ch03fn2a (Accessed: February 9, 2023).

Cherdarchuk, J. (2013) *Data looks better naked*, *Darkhorse Analytics | Edmonton, AB*. Available at: https://www.darkhorseanalytics.com/blog/data-looks-better-naked (Accessed: February 8, 2023).

Plotly-Pattern-matching callbacks (no date) *Plotly*. Dash Python. Available at: https://dash.plotly.com/pattern-matching-callbacks (Accessed: February 12, 2023).

The Data Visualisation Catalogue-Violin plot (no date) *Violin Plot* . The Data Visualisation Catalogue. Available at: https://datavizcatalogue.com/methods/violin_plot.html (Accessed: February 10, 2023).

Tufte, E.R., 2001. The visual display of quantitative information. Available at: http://www.econ.upf.edu/~michael/visualdata/tufte-aesthetics_and_technique.pdf (Accessed: February 8, 2023).