UNIVERSITY OF VICTORIA

CSC 511 / CSC 411 - Information Visualization

Assignment - Storytelling

- Answer all questions on this document directly. Type in your answers respecting the format of the document (10pt Arial) and not removing any existing content; be concise.
- The page limit is **10 pages** including instructions. Do not modify the first two pages/format/headings.
- Upload your completed assignment as a **Word** or **PDF** file on Brightspace.

Background

It is tempting to think of data and visualization as a neutral actor, with a single "correct" set of design choices that "truthfully" report the data. However, outside of egregious errors (e.g., when dates are sorted incorrectly or the y-axis is not scaled uniformly), we see that "ground truth" in data is contextual and situated. Design choices we make give visualization a *rhetorical power* that influences what a reader concludes and remembers about the data, and blurs the line between persuasion and deception. In this exam, you will grapple with these ethical concerns by using visualization to offer two opposing perspectives about a dataset. Make sure you integrate storytelling elements when completing your work.

Your task

You are asked to visualize a dataset about salaries of employees at the University of Victoria from two perspectives.

Step 1: Download the data directly from the Assignment on Brightspace.

Step 2: Start by exploring and familiarizing yourself with the dataset - you are free to use any tool you like such as Tableau or Excel.

Step 3: Once you have developed an understanding of the data, devise a proposition about it: a statement that asserts a judgment or opinion about the trends you might have uncovered. An example of a proposition might be *"The gender pay gap at UVic is increasing."* Write down your proposition at the top of the page for the affirmative visualization and the negation of your proposition at the top of the page for the negative visualization.

Step 4: Design two visualizations to persuade a reader about each side of the proposition: the **affirmative** visualization should persuade the reader that the proposition is true while the **negative** visualization should persuade the reader that the proposition is not true.

- We encourage you to use any design choice that you find produces the most persuasive visualization for each side of your proposition. You are welcome to use techniques we would usually consider to be earnest (e.g., effective and expressive encodings, transparently communicating data transformations, citing sources, etc.) but also those we might sometimes consider to be deceptive (e.g., violating conventions, skewed or slanted titles and labels, truncated scales and axes, filtering outliers, etc.).
- This goal of persuasion also means that, if you use any deception techniques, they should not be immediately obvious as they might otherwise backfire and dissuade your reader.
- You can use any visualization tool to create the visualizations, and choose to sketch by hand as well. You should carefully consider data transformation, visual encoding, textual content (i.e., titles, axes, labels), and annotations. Moreover, in this assignment, we construe "visualization" broadly (i.e., a single visualization can include several concatenated or inset charts).

- The visualizations do not have to show the full dataset (i.e., you may choose a subset of the data to visualize) but still show several dimensions or aggregations of dimensions. The two visualizations can show different transformations or subsets of the data.
- Each visualization must be standalone and clearly convey a story about the data.
- The visualizations must be very different visually (e.g., different layout, encodings, etc..).
- The visualizations must be static, i.e. they do not include interactive features.
- It is not important to encode the exact values (draw sketches, not final visualizations), however, the visualization should faithfully represent the data (i.e. do not make up arbitrary values). In other words, it has to be roughly accurate.

Step 5: Document your design decisions and rationale. For each visualization, list **5 design decisions** you think are central to making the visualization persuasive (note, a "design" decision can also refer to decisions about data transformation or textual content). For each decision:

- **Describe** the design decisions (1 or 2 sentences). Such decisions include (but are not limited to!) data cases and data dimensions, visual variables, data and visual transformations, layout, color, typography, application of design guidelines, consideration of preattentive features, resolution of visual variables, expressiveness and effectiveness of visual variables, etc.. Particular attention must be paid to integrating and describing the use of storytelling/narrative elements.
- Score the decision by **bolding** if it is fully deceptive, deceptive, neutral, earnest, or fully earnest.
- Provide **rationale** that supports the decision. How does this decision help make your visualization persuasive? What other alternatives did you consider, and why did you select this one?
- Note that you could talk about similar design decisions (e.g., you could talk about the title) in both visualizations, however, this should only occur when these are key decisions for both visualizations.

Step 6: Write a final paragraph (~200-300 words) reflecting on your overall design process. What was straightforward or difficult? What surprised you? How do you now define "ethical analysis and visualization"? What bounds (if any) can you draw to distinguish "acceptable" persuasive choices vs. "misleading" ones (and if none, why not)?

Data

The data you are working with to create these two visualizations is the salary information of employees at the University of Victoria between 2013 and 2023. It contains the following fields:
`First_name`: raw first name; `Last_name`: raw last name; `Full_name`: concatenation of First_name and Last_name; `Simplify_name`: simplified first name; `Proba_female`: probability that Simplify_name is female; `Proba_male`: probability that Simplify_name is male; `Department`: home department of the person, or N/A if unknown; `Faculty`: home faculty of the person, or N/A if unknown; `2013` to `2023`: yearly salary of the person in CAD

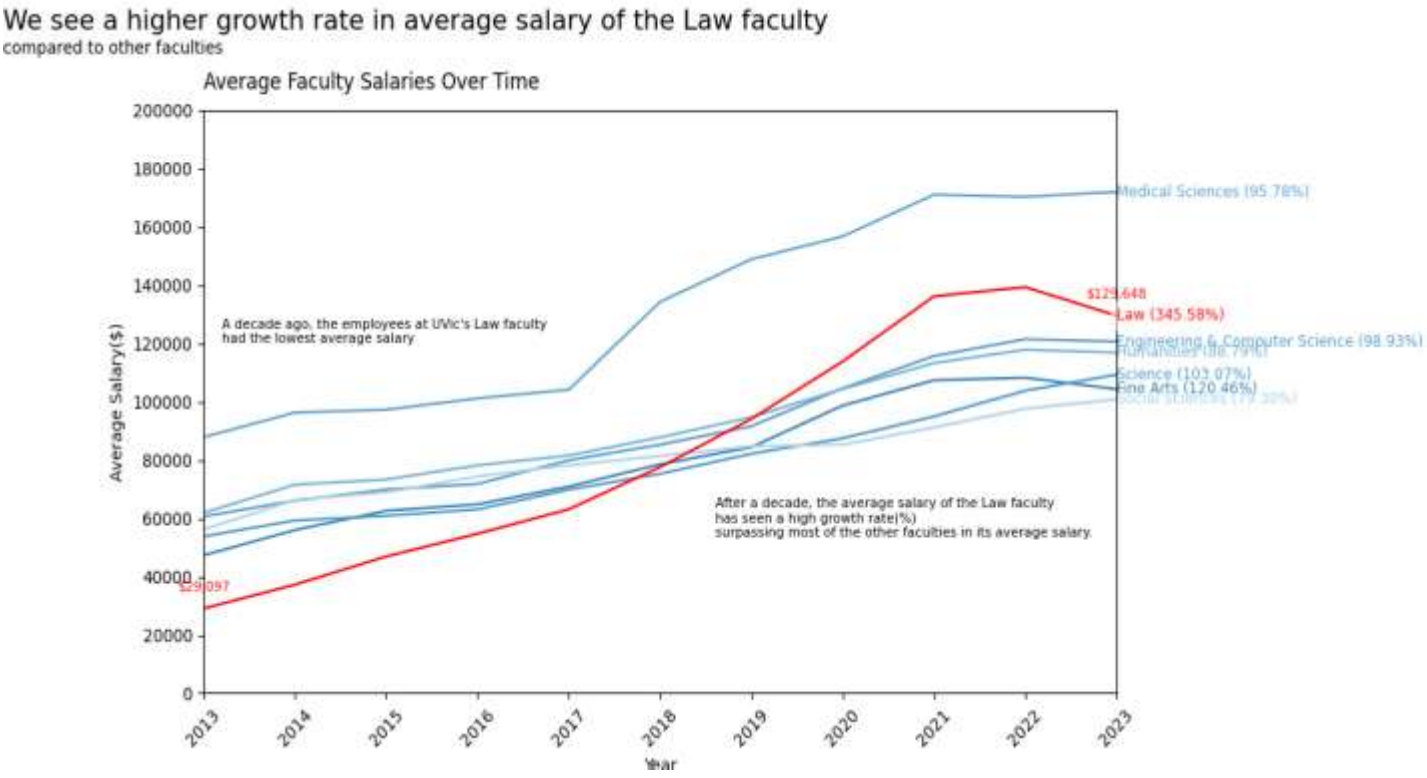
Sources: the yearly salary data was extracted from the University of Victoria's financial statements. Department/Faculty was extracted from this page. Gender probability was computed using the gender R package.

1. Affirmative Visualization [45 marks]

Proposition: The Law faculty at UVic has experienced a significantly higher growth rate in average salaries compared to other faculties.

1.1. Visualization

Paste here an image of your affirmative visualization. This page can only contain that image.



1.2. Design decisions and rationale

Describe 5 design decisions and rationales/justifications.

Design decision (1-2 sentences)	Use of the title: “We see a higher growth rate in average salary of the Law faculty compared to other faculties" directly states the proposition.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	The title directly sets the narrative by clearly stating the proposition, guiding viewers to focus on the specific comparison between the average salary growth rate of the Law faculty and other faculties. This framing is crucial because it primes the viewer to interpret the data with the context of the Law faculty’s significant salary growth. An effective title is a key element in storytelling as it immediately informs the viewer of the main message. In this case, the title succinctly conveys that the Law faculty’s salary growth is noteworthy, leading the audience to look for supporting evidence in the visualization. By doing so, the title helps to establish a narrative focus and ensures that the viewer's attention is directed towards the intended argument, reinforcing the visualization’s persuasive goal. I did initially plan to use a title that just stated, “Average salary of faculties in the past decade” and not just frame my title according to my proposition so that it didn’t seem deceptive. But since the data was in line with my proposition, I considered it to not be a problem and just made my initial title into a subtitle.				

Design decision (1-2 sentences)	Color hue and color value coding: I have used Red for the Law faculty and different Blue color values for the other faculties.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	Red is used to highlight the Law faculty, drawing attention to its trend line, while blue represents other faculties. This distinction ensures the Law faculty's salary growth stands out, reinforcing the narrative. The contrasting colors help viewers easily differentiate between the faculties, making the visualization more effective and the story more compelling. The choice of colors does not distort the data but helps in clearly distinguishing the Law faculty from the rest, making it an earnest decision. The different color value of blue for the other faculties also represent the amount of growth rate(%). This in turn highlights how high the growth rate for the Law faculty is in comparison to the other faculties. One of my initial ideas was to shade the area under the initial and final salary points of the Law faculty and maybe some of the other faculties that had a higher growth rate. But other than the visualization looking too clustered, I also found out that the area doesn’t completely represent the growth rate changes over the years, and hence didn’t implement this design.				

Design decision (1-2 sentences)	Salary Annotation at the beginning and end of the Law faculty average salary trend Line (2013-2023).				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest

Design rationale (150-200 words)	Annotating the Law faculty's salary at key points—$29,097 in 2013 and $129,648 in 2023—emphasizes the dramatic increase over the decade. These annotations serve as clear reference points, highlighting the substantial growth and helping viewers quickly grasp the extent of the change. This design choice supports the narrative by providing concrete data points that illustrate the Law faculty's significant salary increase. It also aids in storytelling by marking critical moments in the data, allowing viewers to follow the progression of salaries over time. By visually anchoring the start and end points of the trend, these annotations make the growth trajectory more relatable and easier to understand. This choice enhances the viewer's comprehension and reinforces the narrative that the Law faculty has experienced a notable salary increase, thereby strengthening the overall argument of the visualization. I didn't implement these annotations for the faculties since I only wanted to emphasize the main comparison factor for the proposition, that is the Law faculty. The others can be interpreted through the position variables.

Design decision (1-2 sentences)	Including growth rate(%) next to each faculty annotations at right side of the graph, i.e. the end of the trend lines for each faculty.				
Score (**bold** answer)	Fully deceptive	Deceptive	**Neutral**	Earnest	Fully earnest
Design rationale (150-200 words)	Adding growth percentages next to each faculty in the legend (e.g., "Law (345.58%)") quantifies the rate of salary increase, making comparisons straightforward for the viewer. This design choice provides an additional layer of context and reinforces the narrative by presenting the data in a clear, understandable manner. By highlighting the different faculty's growth percentage, the visualization draws attention to the magnitude of the Law faculties salary increase relative to other faculties. This choice is effective because it succinctly communicates the rate of growth without cluttering the main chart, allowing viewers to quickly compare the Law faculty's performance with others. The growth percentages serve as a quick reference, making it easier for viewers to grasp the significance of the data at a glance without much bias. This approach enhances the viewer's ability to understand and appreciate the differences in salary growth, thereby supporting the argument that the Law faculty has seen a significantly higher growth rate.				

Design decision (1-2 sentences)	Contextual annotations about the initial and final average salary of the Law faculty and its comparison to the other faculty average salaries.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	The contextual annotations, such as noting the Law faculty's initial low salary and its subsequent high growth rate, add a storytelling element to the visualization. These annotations help viewers understand the background and significance of the data, making the narrative more engaging and persuasive while maintaining transparency. For example, comparing the Law faculty's growth from the lowest average salary in 2013 to surpassing many other faculties by 2023 highlights the remarkable nature of this change. This storytelling approach not only draws attention to the dramatic improvement but also situates it within the broader context of faculty salary trends at UVic. It				

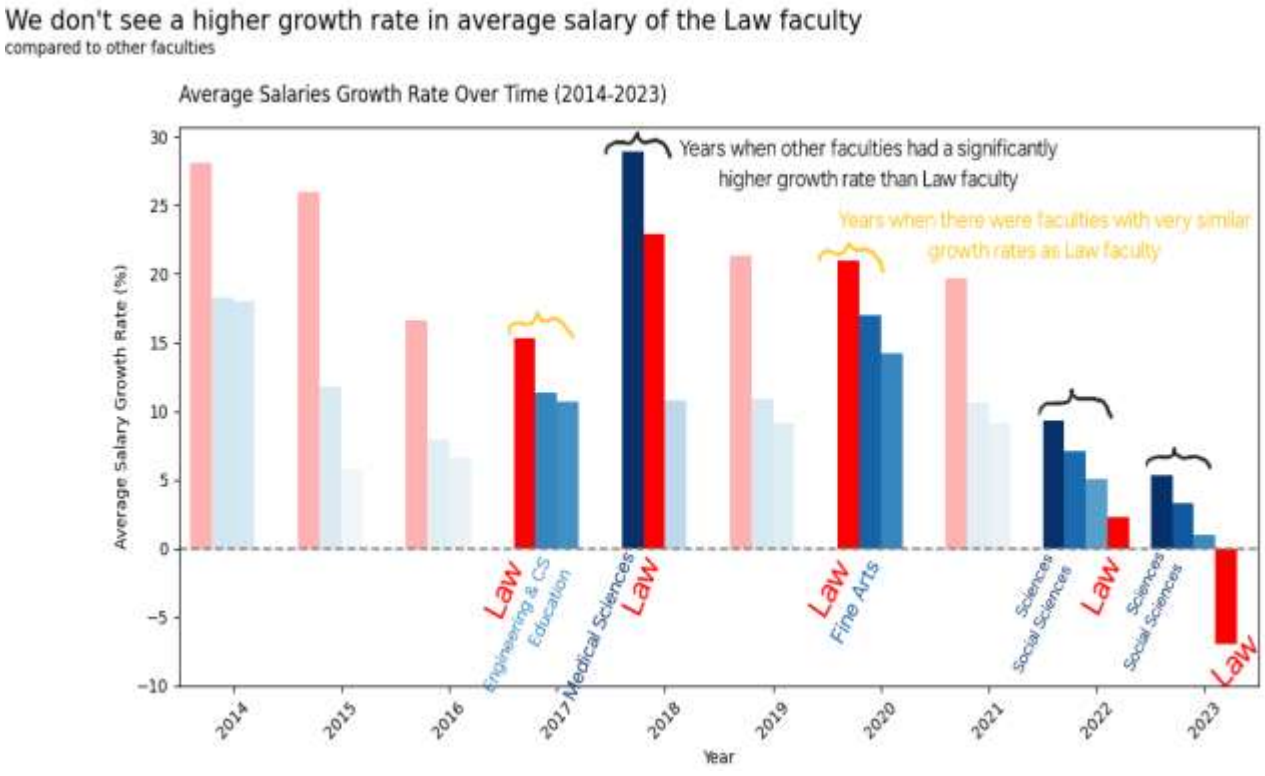
	underscores the narrative that the Law faculty has seen a significant transformation, helping viewers appreciate the scale and impact of the growth in a compelling and contextualized manner. Initially, with my previous idea of shading the area under the lines to represent the growth rates, the contextual annotations would have also been a bit different, emphasizing on the area shaded. But since I scrapped that design idea, I went with the above final design which is relatively much clearer and more concise.

2. Negative Visualization [45 marks]

Negation of proposition: The Law faculty at UVic has not experienced a significantly higher growth rate in average salaries compared to other faculties.

2.1. Visualization

Paste here an image of your negative visualization. This page can only contain that image.



2.2. Design decisions and rationale

Describe 5 design decisions and rationales/justifications.

Design decision (1-2 sentences)	Use of title: "We don't see a higher growth rate in average salary of the Law faculty compared to other faculties."				
Score (**bold** answer)	Fully deceptive	**Deceptive**	Neutral	Earnest	Fully earnest
Design rationale (150-200 words)	Although the title is straightforward and clearly conveys the message that the Law faculty's salary growth rate is not higher compared to other faculties, it is designed to push the narrative that the Law faculty's salary growth rate is not significantly higher, even though the data may show some years where the Law faculty's growth rate was notable. This decision frames the viewer's interpretation from the outset, potentially leading them to overlook instances where the Law faculty's growth was higher. By setting this narrative, the title biases the viewer's perception before they even examine the data. While it is not an outright falsehood, it selectively emphasizes one interpretation of the data, which can be considered deceptive as it influences the reader's understanding by framing the story in a specific way. The other topic I thought of was: "The Law faculty has a lower growth rate than the other faculties", but I realized that rather than negation, it was outright reversing the proposition and hence wrong and fully deceptive.				

Design decision (1-2 sentences)	Highlighting specific years growth rate bars while de-emphasizing some years.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	Highlighting specific years growth rates (emphasized bars) where faculties other than the Law faculty had either significantly higher or very similar growth rates compared to it helps in guiding the reader's attention to critical points in the timeline. This decision ensures that viewers notice the fluctuations and comparisons over time without distorting the data. By highlighting these years bars, the visualization provides context and additional narrative, making it easier for viewers to follow the story being told. This choice helps in emphasizing the key points of comparison, which is central to the negation proposition. While de-emphasized, the other year years are also visible and can show that while the growth rate for the Law faculty is indeed increasing, it is not significantly doing so throughout the decade.				

Design decision (1-2 sentences)	Use of color coding (Red for Law and Blue for other faculties) and color value of Blue.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	Using different colors to represent different faculties, with a consistent color (red) for the Law faculty, helps in quickly identifying and differentiating the Law faculty from others. This aids in making the chart visually distinct and easier to				

	interpret without misleading the viewer. Color coding is a powerful tool in data visualization that, when used ethically, enhances clarity and comprehension. The decision to use color in an earnest manner supports the viewer in making accurate comparisons, reinforcing the overall narrative without introducing any bias. The different gradients of blue for the faculties other than Law also emphasize the varying growth rate of each faculty that year. I also had an idea to include the Law faculty in the range of color value changes according to that year's growth rate but found that emphasizing the main comparison factor, i.e. the Law faculty became much more difficult when visualizing. Hence I left it as a single red color value.

Design decision (1-2 sentences)	Using annotations (colored curly brackets) and context descriptions.				
Score (**bold** answer)	Fully deceptive	Deceptive	Neutral	**Earnest**	Fully earnest
Design rationale (150-200 words)	Adding annotations to explain years when the Law faculty had similar or lower growth rates compared to other faculties provides context and narrative to the data. The yellow curly brackets represent years when other faculties had similar or very small differences in average salary growth rate compared to the Law faculty while the black curly brackets represent the years when other faculties had a higher average salary growth rate than the Law faculty. This design choice supports the viewer in understanding the trends and comparisons without manipulating the data, making the visualization more informative. Annotations act as a guiding tool, helping viewers to focus on important aspects of the data while ensuring that the interpretation remains accurate and fair. This choice enhances the storytelling aspect of the visualization, making it more engaging and insightful. I did want to add growth rate (%) in the bars themselves for quicker viewing but my visualization uses clustered bar graphs and have very thin bars due to lack of space. Due to this when the percentage are included, it looks very messy and difficult to see. So, for clarity purposes, I excluded them from my final design.				

Design decision (1-2 sentences)	Use of position visual variables: Axis Scaling and Labeling.				
Score (**bold** answer)	Fully deceptive	Deceptive	**Neutral**	Earnest	Fully earnest
Design rationale (150-200 words)	The y-axis is uniformly scaled to represent the average salary growth rate, ensuring that the visual comparison between different faculties is fair and accurate. Whereas the x-axis represents year 2014-2023 linearly since growth rate per year can only be calculated onwards from 2014 using 2013 values in its calculation. The labels are clear and precise, aiding in the correct interpretation of the data. Uniform scaling is crucial in maintaining the integrity of the data representation, preventing any distortion that could mislead the viewer. This decision upholds the ethical standards of data visualization by providing a true and fair view of the data, allowing viewers to draw informed conclusions.				

3. Reflection [10 marks]

In ~200-300 words, reflect on your overall design process. What was straightforward or difficult? What surprised you? How do you now define "ethical analysis and visualization"? What bounds (if any) can you draw to distinguish "acceptable" persuasive choices vs. "misleading" ones (and if none, why not)?

The design process for creating visualizations of UVic employee salaries based on an affirmative and a negation proposition was both challenging and enlightening. Initially, coming up with a unique proposition, distinct from provided examples, was quite challenging, especially since the gender dimension seemed important. However, once I settled on a unique proposition, identifying key trends, such as growth rates of average salaries across faculties, became straightforward.

I used Python and the Pandas library for data transformation, calculating average salaries and growth rates. Initially, I sketched my ideas by hand and was planning to complete the whole design process through hand-sketching. However, I found conveying certain design choices, like downplaying or transparency of bars in the negative proposition, difficult. After several iterations, I chose the Matplotlib library, utilising the transformed data (Average salary of each faculty per year and average salary growth rate of each faculty per year) from my Python code to implement the visualizations. Surprisingly, Matplotlib had multiple customization functions that allowed me to implement storytelling elements effectively.

This assignment has refined my understanding of "ethical analysis and visualization" as presenting data truthfully and transparently while engaging the audience. Ethical visualization respects data integrity and audience intelligence, avoiding distortions that could lead to incorrect conclusions. The line between "acceptable" persuasive choices and "misleading" ones lies in the intent and impact of each design decision. Acceptable choices enhance understanding and highlight genuine trends without altering data integrity. In contrast, misleading choices manipulate visuals to create false impressions. The goal should always be to inform and persuade based on actual data, maintaining honesty and clarity.