Interactive features play a crucial role in visual analysis of big datasets, helping users to derive insights from complex and voluminous data. Two of the most important interactive features are:

1. **Dynamic Filtering and Selection:**
   * **Description:** Dynamic filtering and selection allow users to interactively modify what data is visible in a visualization based on certain criteria or selections. This can involve sliders, checkboxes, dropdown menus, or brush tools for selecting ranges of data directly on the visualization itself. For example, in a large dataset containing sales data over several years, dynamic filtering could enable a user to focus on data from a specific year, region, or product category.
   * **Importance:** This feature is critical because it enables users to drill down into large datasets and focus on subsets that are of particular interest, making it easier to uncover patterns, trends, and outliers. It supports exploratory data analysis by allowing users to form and test hypotheses about the data, iteratively refining their questions based on the visual feedback the data provides.
2. **Interactive Dashboards and Linked Visualizations:**
   * **Description:** Interactive dashboards and linked visualizations refer to the coordinated use of multiple data visualizations that update in response to user interactions. When a user interacts with one visualization, such as selecting a segment of a pie chart or a range of values in a histogram, other visualizations on the dashboard update to reflect this selection, showing only data related to the user's interaction.
   * **Importance:** This feature is crucial for analyzing big datasets because it provides a holistic view of the data from multiple perspectives simultaneously. It allows for cross-filtering, where interacting with one element of the dashboard dynamically influences the content of others, facilitating a deeper understanding of the relationships and correlations within the data. This multi-faceted approach to data exploration helps in identifying patterns and insights that might not be apparent from a single visualization, making it an essential tool in the analysis of complex datasets.

These features are integral to the visual analysis of big datasets because they empower users to navigate and make sense of the complexity and scale of the data. By enabling interactive exploration, users can more effectively identify relevant patterns, trends, and insights, leading to more informed decision-making and analysis.

At university X, the late graduation rate is high meaning that most students cannot finish the program with the designed timeline of the curriculum.

The reasons could be that:

* Students cannot pass courses after finish attempt,
* Students did not register for courses in the designated semester,
* Students intentionally lengthen the leaning path by registering for fewer courses than the suggested courses in the curriculum,
* Or the reasons

Hence, people would like to analyze the learning progress of students to make early warnings to students and the dean of the school regarding the potential late graduation problem.

* The data for the learning progress of student is grouped by enrolment batch.
* Each patch has an enrolment year and a major i.e. Batch 2020-Computer Science, Batch 2021-Data Science, and Batch 2019-Information Technology
* Each batch has it curriculum inclosing a list of courses students of the batch should take each semester like the following table:

+ Batch – Semester – Course

+ 2020-CS – 1 - Introduction to computing (3 credits), C++ Programing (4 credits)

+ 2020-CS – 2 – OOP (4 credits), CN (4 credits)

* Each bath has a summarized of grades of students by semester like the following table:

+ Batch – Semester/Year – Course – Number of registrations/Total students of the batch – Number of students who passed the course (score > 50)

+ 2020-CS – Sem1/2020 – Introduction to Computing – 60/60 – 60

+ 2020-CS – Sem1/2020 – C/C++ Programming – 50/60 – 40

+ 2020-CS – Sem1/2020 – C/C++ Programming – 5/60 – 5

+ 2020-CS – Sem2/2020 – CN – 60/60 – 50

* In the table above, row 2 shows that 50 students of 60 students of batch 2020-Computer Science registered for “C/C++ Programming” in semester 1 of 2020. Add only 40 of those 50 students passed the course.
* Row 3 show that in semester 2 of 2020, 5 students of batch 2020-Computer Science registerd for “C/C++ Programming” and all those 5 students passed the course?

Your tasks are to design chart(s) to analyze the learning progress of students and highlight potential late graduation rate.

1. (10pts) State the problem described above in terms of abstract task in visualization.
2. (10pts) What are types of data? What are the marks and channels to use? Draw the sample chart(s) to illustrates your design.
3. (10pts) Explain and describe how your chart(s) could help people detect the potential high late graduation r

Interactive features play a crucial role in visual analysis of big datasets, helping users to derive insights from complex and voluminous data. Two of the most important interactive features are:

1. **Dynamic Filtering and Selection:**
   * **Description:** Dynamic filtering and selection allow users to interactively modify what data is visible in a visualization based on certain criteria or selections. This can involve sliders, checkboxes, dropdown menus, or brush tools for selecting ranges of data directly on the visualization itself. For example, in a large dataset containing sales data over several years, dynamic filtering could enable a user to focus on data from a specific year, region, or product category.
   * **Importance:** This feature is critical because it enables users to drill down into large datasets and focus on subsets that are of particular interest, making it easier to uncover patterns, trends, and outliers. It supports exploratory data analysis by allowing users to form and test hypotheses about the data, iteratively refining their questions based on the visual feedback the data provides.
2. **Interactive Dashboards and Linked Visualizations:**
   * **Description:** Interactive dashboards and linked visualizations refer to the coordinated use of multiple data visualizations that update in response to user interactions. When a user interacts with one visualization, such as selecting a segment of a pie chart or a range of values in a histogram, other visualizations on the dashboard update to reflect this selection, showing only data related to the user's interaction.
   * **Importance:** This feature is crucial for analyzing big datasets because it provides a holistic view of the data from multiple perspectives simultaneously. It allows for cross-filtering, where interacting with one element of the dashboard dynamically influences the content of others, facilitating a deeper understanding of the relationships and correlations within the data. This multi-faceted approach to data exploration helps in identifying patterns and insights that might not be apparent from a single visualization, making it an essential tool in the analysis of complex datasets.

These features are integral to the visual analysis of big datasets because they empower users to navigate and make sense of the complexity and scale of the data. By enabling interactive exploration, users can more effectively identify relevant patterns, trends, and insights, leading to more informed decision-making and analysis.



**You**

At university X, the late graduation rate is high meaning that most students cannot finish the program with the designed timeline of the curriculum. The reasons could be that: - Students cannot pass courses after finish attempt, - Students did not register for courses in the designated semester, - Students intentionally lengthen the leaning path by registering for fewer courses than the suggested courses in the curriculum, - Or the reasons Hence, people would like to analyze the learning progress of students to make early warnings to students and the dean of the school regarding the potential late graduation problem. - The data for the learning progress of student is grouped by enrolment batch. - Each patch has an enrolment year and a major i.e. Batch 2020-Computer Science, Batch 2021-Data Science, and Batch 2019-Information Technology - Each batch has it curriculum inclosing a list of courses students of the batch should take each semester like the following table: + Batch – Semester – Course + 2020-CS – 1 - Introduction to computing (3 credits), C++ Programing (4 credits) + 2020-CS – 2 – OOP (4 credits), CN (4 credits) - Each bath has a summarized of grades of students by semester like the following table: + Batch – Semester/Year – Course – Number of registrations/Total students of the batch – Number of students who passed the course (score > 50) + 2020-CS – Sem1/2020 – Introduction to Computing – 60/60 – 60 + 2020-CS – Sem1/2020 – C/C++ Programming – 50/60 – 40 + 2020-CS – Sem1/2020 – C/C++ Programming – 5/60 – 5 + 2020-CS – Sem2/2020 – CN – 60/60 – 50 - In the table above, row 2 shows that 50 students of 60 students of batch 2020-Computer Science registered for “C/C++ Programming” in semester 1 of 2020. Add only 40 of those 50 students passed the course. - Row 3 show that in semester 2 of 2020, 5 students of batch 2020-Computer Science registerd for “C/C++ Programming” and all those 5 students passed the course? Your tasks are to design chart(s) to analyze the learning progress of students and highlight potential late graduation rate. a) (10pts) State the problem described above in terms of abstract task in visualization. b) (10pts) What are types of data? What are the marks and channels to use? Draw the sample chart(s) to illustrates your design. c) (10pts) Explain and describe how your chart(s) could help people detect the potential high late graduation rate of a batch.

ChatGPT

**ChatGPT**

**a) Abstract Task in Visualization**

The primary task is to design a visualization system that enables the identification and analysis of factors contributing to late graduation rates among university students by batch and major. This system should facilitate the exploration of course registration patterns, pass rates, and adherence to the prescribed curriculum timeline, enabling stakeholders to make informed decisions and intervene where necessary.

**b) Types of Data, Marks, and Channels**

* **Types of Data:**
  1. **Categorical Data:** Batch identification (e.g., "2020-CS"), majors (e.g., "Computer Science"), semesters (e.g., "Sem1/2020"), and courses (e.g., "Introduction to Computing").
  2. **Quantitative Data:** Number of registrations per course, total students per batch, and number of students who passed each course.
* **Marks and Channels:**
  1. **Marks:** Points, lines, bars.
  2. **Channels:** Position, length, color, and size.
* **Sample Chart Design:**
  1. **Stacked Bar Chart for Course Registration and Pass Rates by Semester:**
     + **X-Axis:** Semesters (time-ordered).
     + **Y-Axis:** Number of students.
     + **Bars:** Each bar represents a course, with the length indicating the number of registrations.
     + **Segments within Bars:** Differentiate between students who passed and those who didn't using color (e.g., green for passed, red for not passed).
     + **Faceting:** Separate charts for each batch and major to allow comparison.
  2. **Line Chart for Cumulative Course Completions:**
     + **X-Axis:** Semesters.
     + **Y-Axis:** Cumulative number of courses completed by students.
     + **Lines:** Represent individual students or an average trend line for the batch.
     + **Color:** Differentiate lines by major or other relevant categories.
* **Sample Visualization:**

Let's create a simplified illustration of the stacked bar chart design for course registration and pass rates by semester, which will provide a visual representation of how this concept might be applied to the data described.

dalleCopy code

{ "prompt": "Create an illustrative stacked bar chart showing course registration and pass rates for a university batch. The x-axis should have labels for semesters, such as 'Sem1/2020', 'Sem2/2020', and so on. Each bar represents a course, with the length indicating the number of registrations. Within each bar, use green to show the portion of students who passed and red for those who didn't. Include a legend indicating what green and red represent. Facet the chart by batches and majors, such as '2020-CS', '2021-Data Science', to allow comparison between different groups. Add a title 'Course Registration and Pass Rates by Semester for University Batches'.", "size": "1024x1024" }

**c) Detection of Potential High Late Graduation Rate**

The designed charts would help in detecting potential high late graduation rates in several ways:

1. **Identifying Challenging Courses:**
   * The stacked bar chart can highlight courses with low pass rates, indicating subjects that may be barriers to timely graduation.
2. **Trends in Course Registrations:**
   * By observing the registration numbers across semesters, stakeholders can identify patterns where students might be avoiding certain courses or registering for fewer courses than suggested, potentially delaying graduation.
3. **Comparison Across Batches:**
   * Faceting by batch and major allows for the comparison of course completion rates and potential issues across different groups, helping to identify if certain batches or majors are more prone to late graduations.
4. **Cumulative Progress Tracking:**
   * The line chart showing cumulative course completions provides a clear visual representation of students' progress towards their degree requirements over time. Deviations from the expected trajectory can prompt early interventions.
5. **Actionable Insights:**
   * The visualization can serve as a basis for discussions among educators and administrators about curriculum adjustments, support systems, and interventions to assist students at risk of late graduation.
6. ate of a batch.