EDAV COMMUNITY CONTRIBUTION

GROUP NUMBER: CC5
- **Varalika Mahajan**: vm2695
- **Vrinda Bhat**: vgb2113

Visualization Preference Analysis

November 10, 2022

Overview

- **Form Link:** https://forms.gle/SYKeS6fqGktr5qrN8
- **No of responses: 80**
- **Data Collected:** https://docs.google.com/spreadsheets/d/1apWvWCVmk4donteDaWEuMjDvHc1gJZxfrbSXeILNJyE/edit?usp=sharing

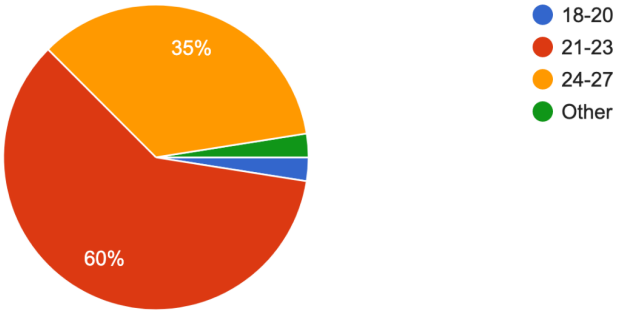
Goals

1. **To understand people's opinions on different visualizations:** Different views on graphs and visualizations help us to understand their importance and the structural necessity of existing and most-used graphs.

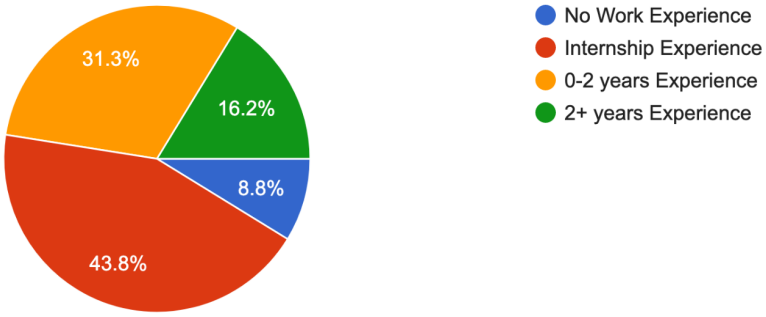
2. **Validate if factors like age, work experience, and gender impact their opinion:** Usually applying the learned classroom concepts to real-life work-related problems gives us a better understanding of what is more meaningful for business insights. Thus, we checked if factors like experience, age, and gender showed a huge impact on choices.

Our Outcomes

Age group (years)
80 responses

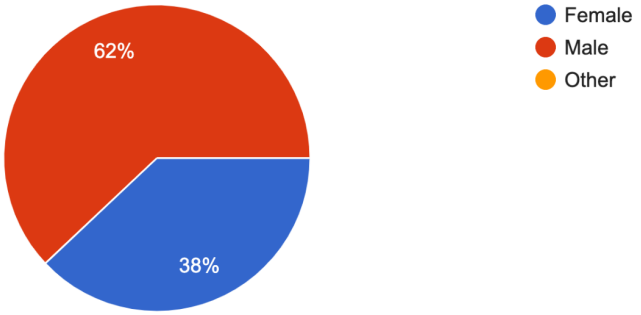


Work Experience
80 responses



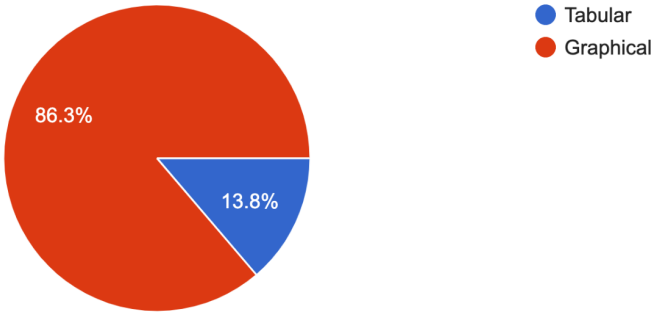
Gender

79 responses



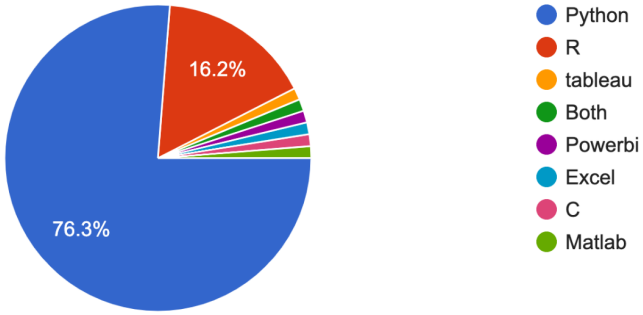
Given this type of data, do you prefer tabular summarization or graphical summarization of the data?

80 responses



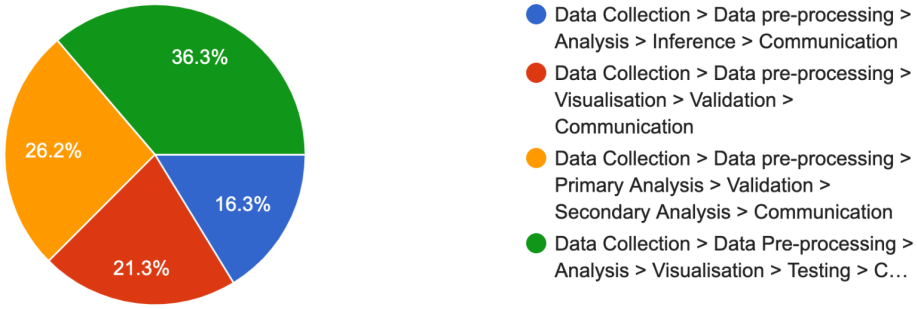
2. Which programming language do you prefer for data analysis or visualization?

80 responses



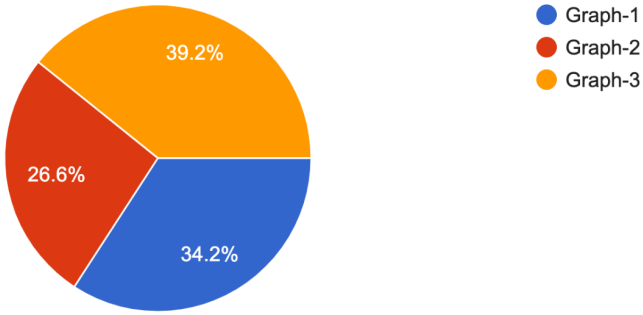
3. Which set of procedures do you prefer while performing exploratory data analysis for any dataset?

80 responses



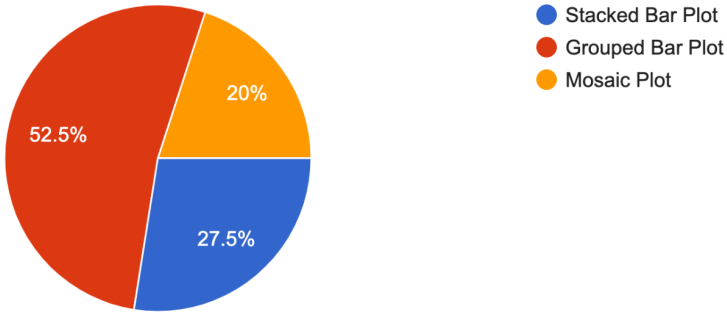
Which dependent variable do you prefer among the below variations of graph, depicting the relation between gender, rank and discipline?

79 responses



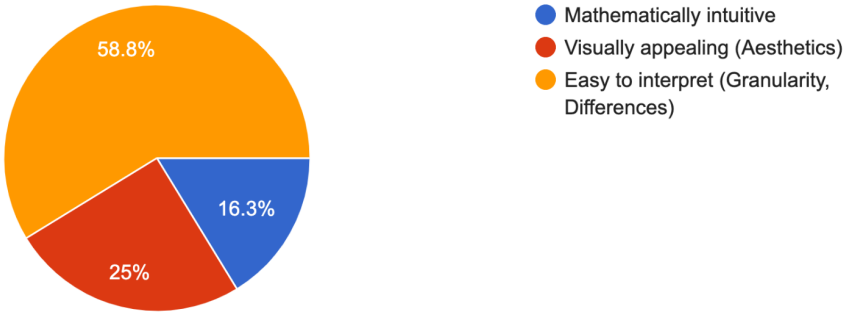
Which type of graph do you prefer to depict the relationship between the following data points: 1. Rank 2. Discipline 3. Gender

80 responses



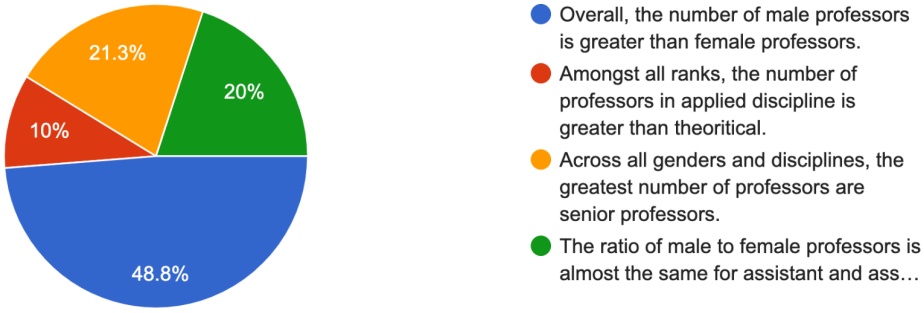
6. Why did you choose this type of graph?

80 responses



7. What is your first observation about the data from the graph?

80 responses



Overall Inference

1. Graphical vs Tabular representation of data

   Graphical representation of data is preferred over tabular summarization by most of students.

   Graphical representation: 86.3%

   Tabular representation: 13.8%

2. Python vs R vs Others

   While the distribution of the preferred language used for exploratory data analysis and visualization varies between students, the most preferred language is Python over R. Other languages such as Matlab and C are also preferred by some students.

   Python: 76.3%

   R: 16.2%

   Others: 7.5%

3. Data Processing techniques

   36% of the users preferred the process: Data Collection > Data Pre-processing > Analysis > Visualisation > Testing > Communication.

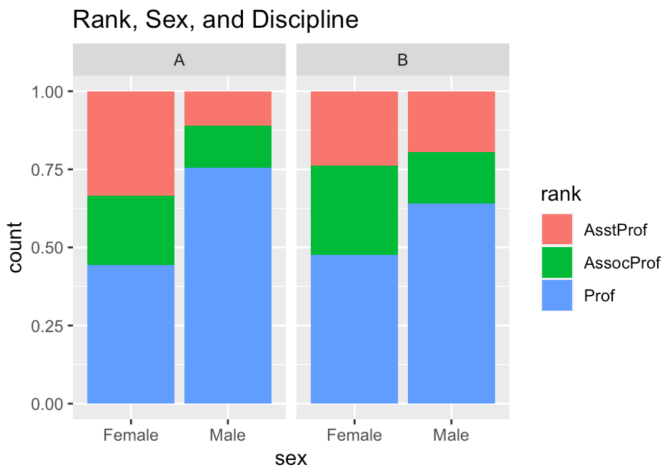
   However, with **26.2%**, Data Collection > Data pre-processing >Primary Analysis > Validation >Secondary Analysis > Communication was the second most preferred.

4. Graph depiction between gender, rank, and description

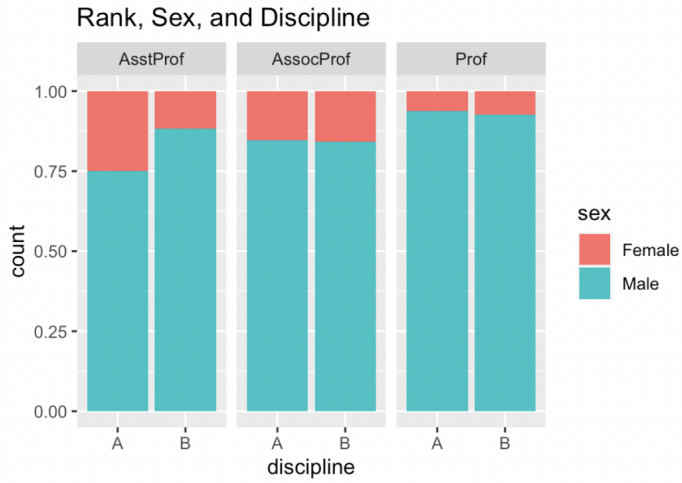
The preferred type of graph is Graph 3(39.2%).

The second most preferred is Graph 1 with 34.2%

Graph-3

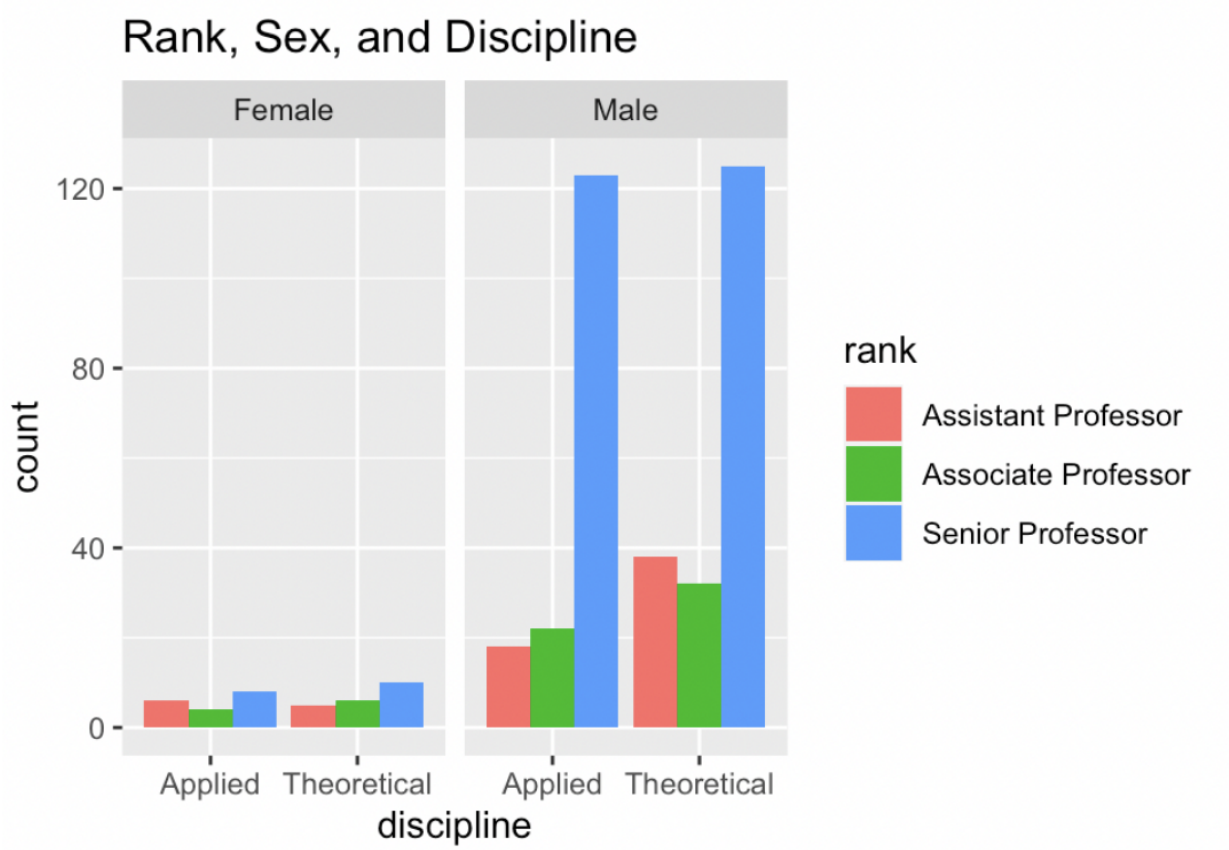


Graph-1



5. Stacked vs Grouped vs Mosaic plots

The preferred type of graph to depict the relationship between gender, rank, and description is grouped bar plots(52.5%). This particular type of graph was chosen as it is easy to interpret based on the granularity and the differences.



6. Primary observation from the graph

The first observation made by students(48.8%) regarding the graph is - Overall, the number of male professors is greater than female professors.

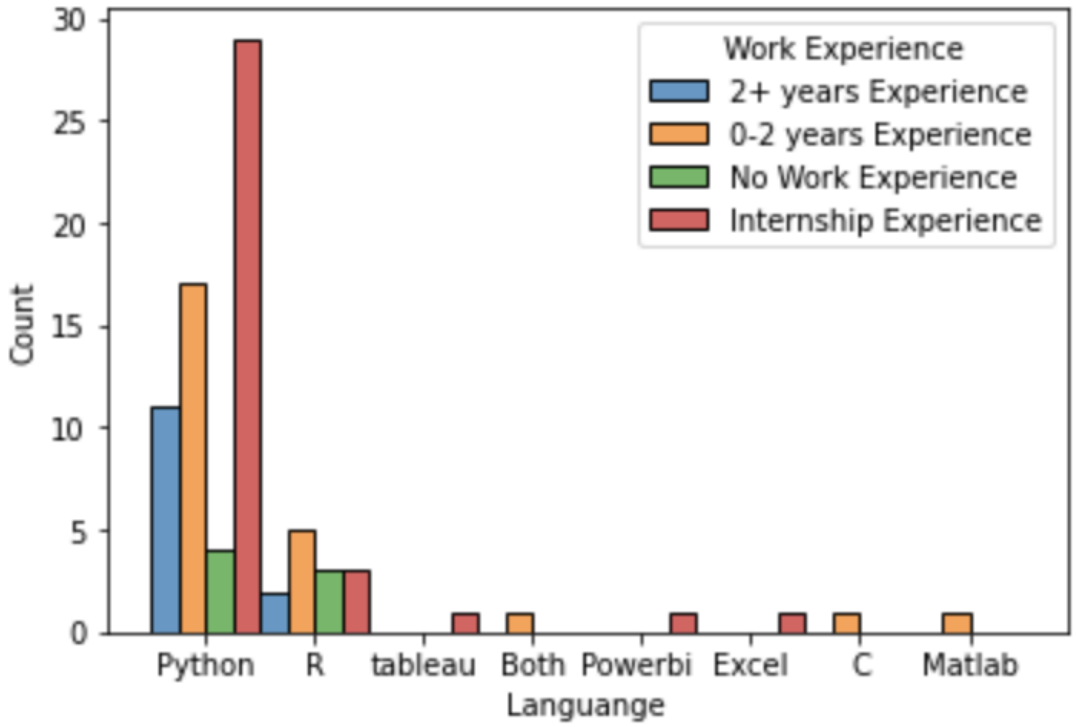
Which is the most observed pattern in the preferred graphs above.

IMPACT ANALYSIS

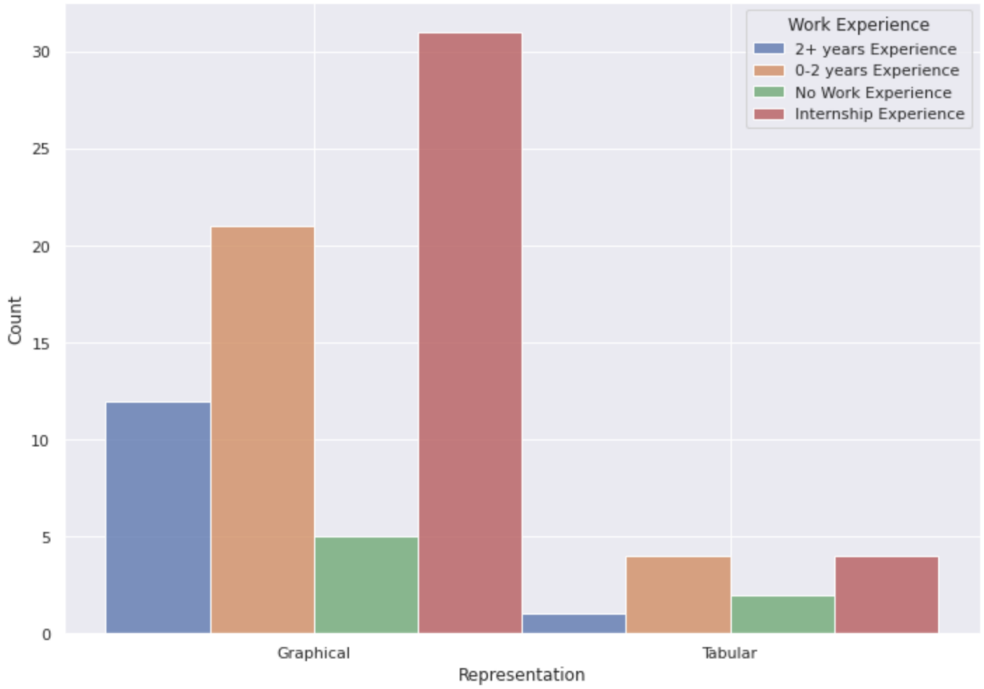
Our dataset contains balanced observations of gender and most people are in the age-group 18-27 years. As a matter of fact, most of them are Columbia graduates pursuing data science and computer science engineering courses. Few students have backgrounds in finance, business, and statistics. The only differentiating factor is work experience.

Impact of Work-Experience on Graph Choices:

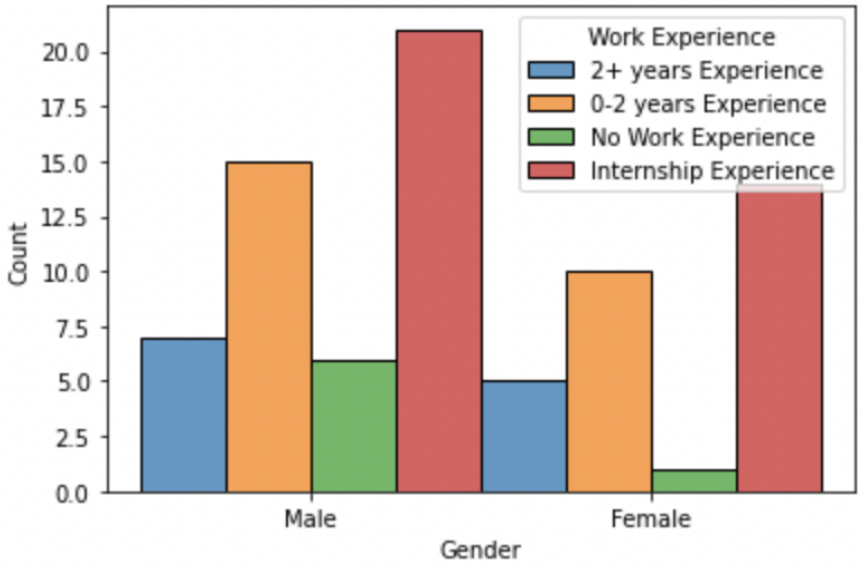
1. **Work-experience vs Language:** The languages preferred for EDAV varies among students. The most preferred languages are Python and R. The other languages which are popularly used are C, Matlab, Excel, Tableau, PowerBI, and others.



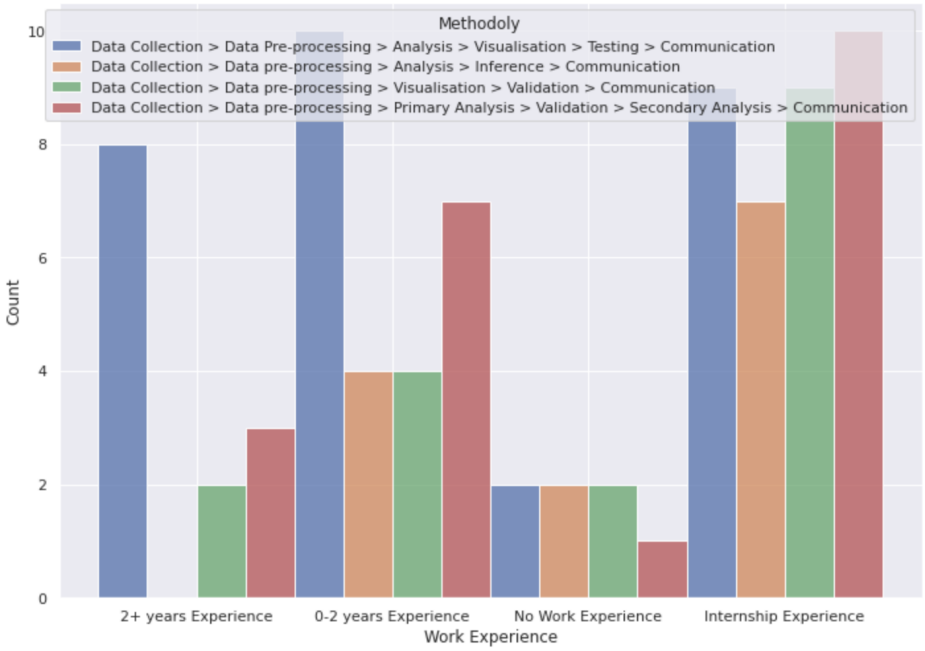
2. **Work experience vs Representation:** Most students prefer a graphical representation of data than a tabular one as it is easier to understand the insights and create visualizations.



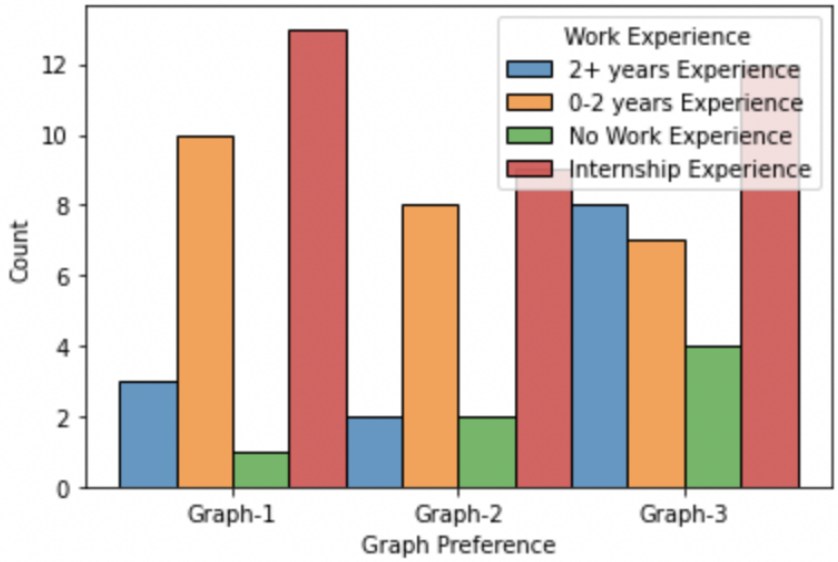
3. **Work-experience vs Gender:** The number of males having internship experience, 0-2 years and 2+ years of experience are higher than that of females



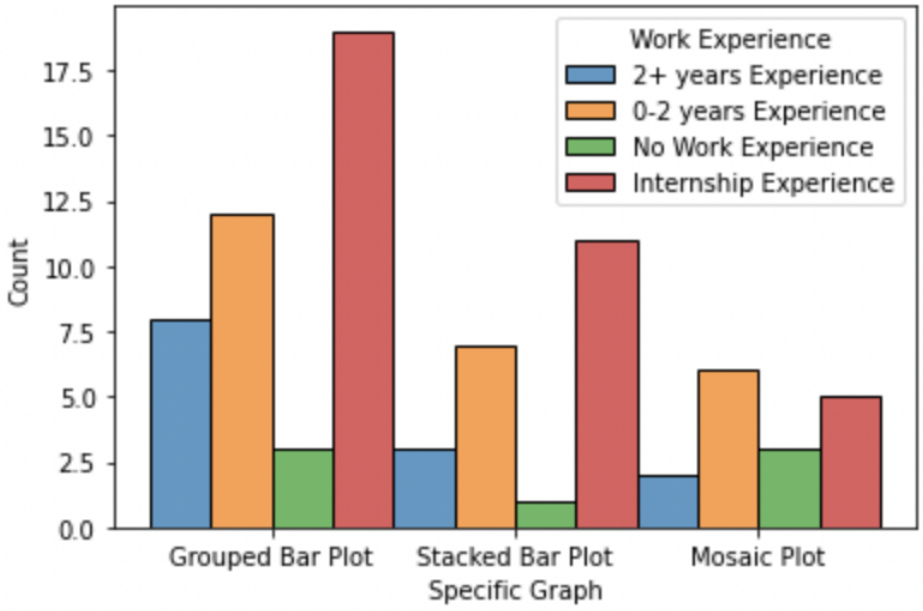
4. **Work experience vs Methodology:** Students prefer analyzing the data before visualizing and then testing to communicate the results. This is seen among students having some experience.



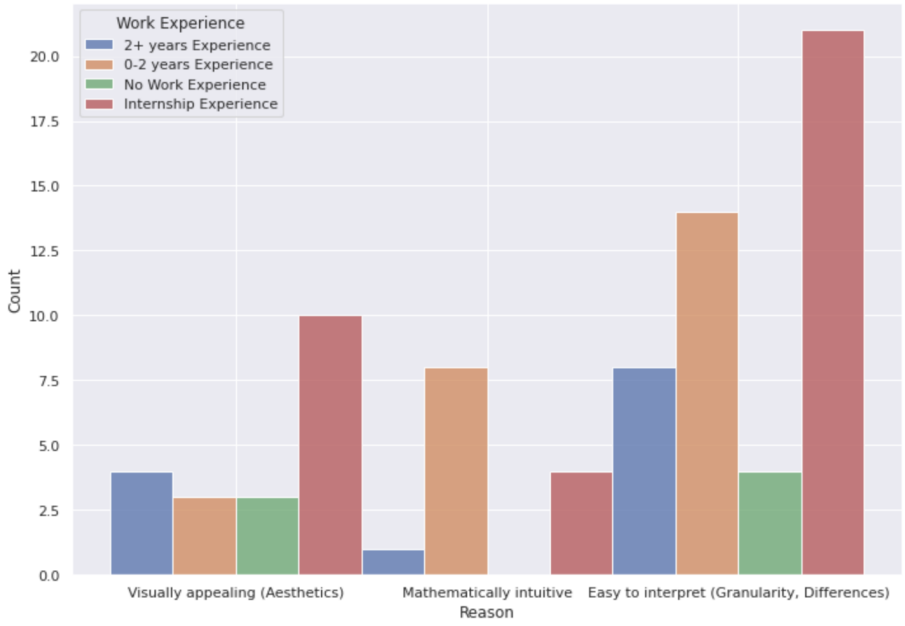
5. **Work experience vs Graph preference:** Graph-3 is the most preferred type of representation for the given parameters.



6. **Work experience vs Specific graph:** The mosaic plot is the least preferred whereas the grouped bar plot is the most preferred type of graph amongst students with and without experience.



7. **Work experience vs Reason:** The preferred type of reasoning to choose the above graph is interpretability. Aesthetics takes the second position amongst graph preference.



8. **Work experience vs First observation:** Amongst all, the first observation of most of the students is that the count of male professors is greater than that of female professors.