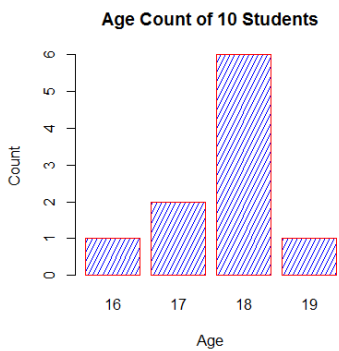


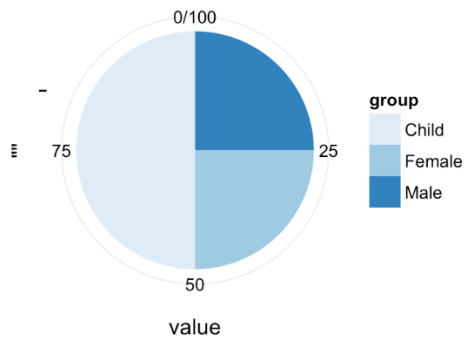
Visualizing and Summarizing Categorical Data

When conducting exploratory data analysis (EDA) on categorical data, the goal is to gain insights into the distribution, frequencies, relationships, and patterns within the categorical variables. Here are some common exploratory analysis techniques used for categorical data:

1. **Frequency Distribution:** Calculate and visualize the frequencies or counts of each category within the categorical variable. This provides an overview of the distribution of the data and helps identify the most common or rare categories.
2. **Bar Plots:** Create bar plots or histograms to visually represent the frequencies of each category. This allows for easy comparison of the relative frequencies and identification of the dominant or minority categories.



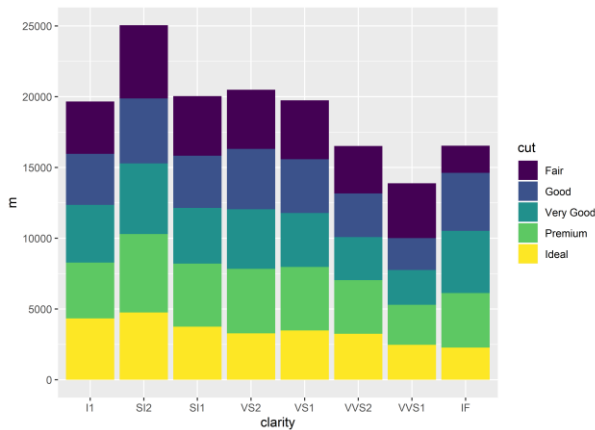
3. **Pie Charts:** Use pie charts to represent the proportion of each category within the categorical variable. Pie charts are useful for understanding the relative contribution of each category to the whole.



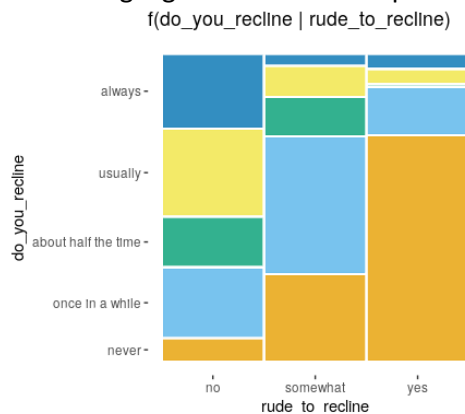
4. **Cross-tabulation (Contingency Table):** Construct cross-tabulations or contingency tables to explore the relationships between two or more categorical variables. This allows you to examine the frequencies and associations between different categories.

<i>Memory</i>	<i>Attention</i>		<i>Total</i>
	Distracted	Focused	
High	10	30	40
Low	47	13	60
<i>Total</i>	57	43	100

5. **Chi-Square Test:** Conduct a chi-square test of independence to determine if there is a statistically significant relationship between two categorical variables. This test helps assess if the observed frequencies differ significantly from the expected frequencies under the assumption of independence.
6. **Mode:** Identify the mode, which represents the category with the highest frequency within a categorical variable. The mode helps to understand the most common category and its dominance in the data.
7. **Association Measures:** Calculate association measures such as chi-square, Cramer's V, or contingency coefficients to quantify the strength and direction of the association between two categorical variables.
8. **Stacked Bar Plots:** Use stacked bar plots to visualize the distribution of a categorical variable broken down by another categorical variable. This allows for the comparison of category proportions within different groups or levels of the other variable.



9. **Mosaic Plots:** Construct mosaic plots to visualize the joint distribution of two categorical variables. Mosaic plots provide a graphical representation of the association between the variables and highlight the relative frequencies within each category.

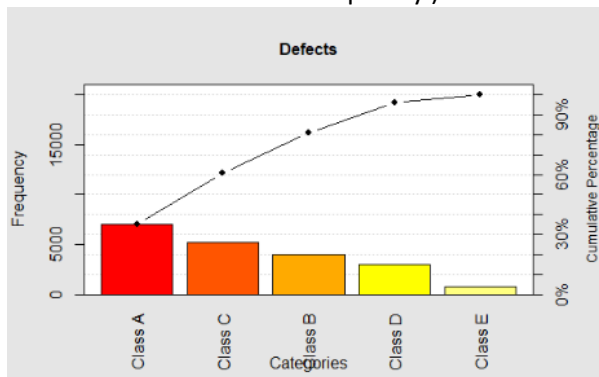


10. **Correspondence Analysis:** Perform correspondence analysis to explore relationships and dependencies among categorical variables. Correspondence analysis helps visualize the patterns and associations in high-dimensional categorical data.

These are some common techniques used in exploratory data analysis for categorical data. The choice of techniques depends on the specific characteristics and goals of the analysis, as well as the number of categorical variables involved. It's important to select appropriate visualizations and statistical measures to uncover insights and patterns within the categorical data.

When conducting exploratory data analysis (EDA) on ordinal data, which consists of variables with ordered categories, the goal is to understand the distribution, central tendency, relationships, and other characteristics of the data. Ordinal data represents a ranking or order, but the numerical difference between categories may not be meaningful or consistent. Here are some common exploratory analysis techniques used for ordinal data:

1. Frequency Distribution: Calculate and visualize the frequencies or counts of each category within the ordinal variable. This provides an overview of the distribution of the data and helps identify the most common or rare categories.
2. Bar Plots: Create bar plots or histograms to visually represent the frequencies of each category. Bar plots help visualize the relative frequencies and identify the dominant or minority categories.
3. Cumulative Frequency Distribution: Construct cumulative frequency distributions to examine the proportion of data falling below each category. Cumulative frequency plots provide insights into the overall distribution and can help identify thresholds or breakpoints. (below is a Pareto chart that includes cumulative frequency.)



4. Descriptive Statistics: Calculate summary statistics such as median, mode, and percentiles. These statistics provide measures of central tendency and spread that are appropriate for ordinal data.
5. Boxplots: Generate boxplots to visualize the distributional properties of the ordinal data. Boxplots display the median, quartiles, and potential outliers, providing a summary of the central tendency and spread of the data.
6. Cross-tabulation: Conduct cross-tabulations or contingency tables to explore the relationships between two or more ordinal variables. Cross-tabulations help analyze the joint distribution of categories and identify potential associations or patterns.
7. Spearman's Rank-Order Correlation: Calculate Spearman's rank-order correlation coefficient to measure the strength and direction of the monotonic relationship between two ordinal variables. Spearman's correlation is appropriate for ordinal data as it captures the ordinal nature of the variables.
8. Ordinal Regression: Apply ordinal regression models, such as the proportional odds model or the cumulative logit model, to analyze the relationship between an ordinal response variable and one or more predictor variables. Ordinal regression helps quantify the effect of predictors on the ordinal outcome.
9. Nonparametric Tests: Use nonparametric tests, such as the Mann-Whitney U test or the Kruskal-Wallis test, to compare ordinal variables across different groups or conditions. Nonparametric tests are appropriate when assumptions of normality or equal variances are not met.
10. Visualizations: Explore visualizations specific to ordinal data, such as parallel coordinate plots or mosaic plots, to reveal relationships and patterns within the data.

These are some common techniques used in exploratory data analysis for ordinal data. The choice of techniques depends on the specific characteristics of the data and the research questions or objectives. It's important to leverage appropriate statistical measures and visualizations to gain insights into the distribution, relationships, and patterns within the ordinal data.

Resources:

1. <https://blog.rsquaredacademy.com/handling-categorical-data-in-r-part-4/>
2. <https://openintro-ims.netlify.app/explore-categorical>
3. <https://rpubs.com/Shairam/734097>
4. <https://dzchilds.github.io/eda-for-bio/exploring-categorical-variables.html>
5. <https://rpubs.com/odenipinedo/exploratory-data-analysis-in-R>
6. <https://campus.datacamp.com/courses/exploratory-data-analysis-in-r/exploring-categorical-data?ex=1>
7. <https://mtstateintrostats.github.io/IntroStatTextbook/explore-categorical.html>
8. <https://www.geeksforgeeks.org/exploring-categorical-data/>
9. <https://www.r-bloggers.com/2019/06/exploratory-data-analysis-with-categorical-data/>
10. <https://www.marsja.se/cross-tabulation-in-r-creating-interpreting-contingency-tables/>
11. <https://www.datamentor.io/r-programming/bar-plot>
12. <https://cran.r-project.org/web/packages/ggmosaic/vignettes/ggmosaic.html>
13. <https://www.geeksforgeeks.org/r-pareto-chart/>