**Visualizations for One and Two Variables**

**Visualizations for One Variable:**

|  |  |  |
| --- | --- | --- |
| Graph Type | Uses | Examples |
| Histogram or density plot | Examine data range  Check number of modes  Check if distribution is normal/lognormal  Check for anomalies and outliers | Examine the distribution of peoples age to get a typical age range.  Examine income levels to get typical income range. |
| Bar chart or dot plot | Compare frequencies of the values of a categorical variable | Count the number of customers from different states of residence to determine which states have the largest or smallest customer base. |

**Visualizations for Two Variables:**

|  |  |  |
| --- | --- | --- |
| Graph Type | Uses | Examples |
| Line Plot | Shows the relationship between two continuous variables.  Best when that relationship is functional, or nearly so. | Plot y = f(x) |
| Scatter Plot | Shows the relationship between two continuous variables.  Best when the relationship is too loose or cloud-like to be easily seen on a line plot. | Plot income vs. years in the workforce (income on y-axis). |
| Smoothing Curve | Shows underlying “average” relationship, or trend, between two continuous variables.  Can also be used to show the relationship between a continuous and a binary or Boolean variable: the fraction of ‘true’ values of the discrete variable as a function of the continuous variable. | Estimate the “average” relationship of income to years in the workforce. |
| Hexbin Plot | Shows the relationship between two continuous variables with the data is very dense. | Plot income vs. years in the workforce for a large population. |
| Stacked Bar Chart | Shows the relationship between two categorical variables (var1 and var2).  Highlights the frequencies of each value of var 1. Works best when var2 is binary. | Plot insurance coverage (var2) as a function of marital status (var1) when you wish to retain information about the number of people in each marital category. |
| Side-by-side Bar Chart | Shows the relationship between two categorical variables (var1 and var2).  Good for comparing the frequencies of each value of var2 across the values of var1. Works best when var2 is binary. | Plot insurance coverage (var2) as a function of marital status (var1) when you wish to directly compare the number of insured and uninsured people in each marital category. |
| Shadow Bar Chart | Shows the relationship between two categorical variables (var1 and var2).  Displays the frequency of each value of var1, while allowing comparison of var2 values both within and across categories. |  |
| Filled Bar Chart | Shows the relationship between two categorical variables (var1 and var2).  Good for comparing the relative frequencies of each value of var2 within each value of var1. Works best when var2 is binary. | Plot insurance coverage (var2) as a function of marital status (var1) when you wish to compare the *ratio* of uninsured to insured people in each marital category. |
| Bar Chart with Faceting | Shows the relationship between two categorical variables (var1 and var2).  Best for comparing the relative frequencies of each value of var2 within each value of var 1 when var2 takes on more than two values | Plot the distribution of marital status (var2) as function of housing type (var1). |
| Overlaid Density Plot | Compares the distribution of a continuous variable over different values of a categorical variable.  Best when the categorical variable has only two or three categories. Shows whether the continuous variable is distributed differently or similarly across the categories. | Compare the age distribution of married vs. divorced populations. |
| Faceted Density Plot | Compares the distribution of a continuous variable over different values of a categorical variable.  Suitable for categorical variables with more than three or so categories. Shows whether the continuous variable is distributed differently or similarly across the categories. | Compare the age distribution of several marital statuses (never married, married, divorced, widowed). |
| Faceted Histogram or shadow histogram | Compares the distribution of a continuous variable over different values of a categorical variable while retaining information about the relative population sizes | Compare the age distribution of several marital statuses while retaining information about relative population sizes. |