An Introduction to Data Visualization and Summary Statistics

Data visualization and summary statistics are an important part of statistical analysis. It can help you identify trends in your data and communicate your research in presentations. Here are some recommendations of plots and descriptive statistics you can use, based on the type of data you have. Examples of all of the charts and graphs mentioned are listed at the end.

**Continuous Numeric Data:**

When your data is a set of numbers between a range. The data can take any value over that interval. If the interval is 1-10, then the data can take values of 2, 3.4, 7.6, .391234 and so on.

Appropriate charts: Histograms, box plots
Appropriate descriptive statistics: Mean, Standard Deviation or Variance, Range (Maximum/Minimum), Median (more appropriate skewed data)

**Nominal Categorical Data:**

Is when your data fits into categories without ranks, for example: 'Red', 'Green', 'Blue', or 'yes'/no'. While the colors or response are different Red is not higher or lower than Green.

Appropriate charts: Bar charts
Appropriate descriptive statistics: Frequency table, mode

**Ordinal Categorical Data:**

This is when you have data that has distinct categories, that have an order to them, like 'low, 'medium', 'high' setting on a machine or '3 months', '4 months', '5 months' as distinct time units. Months could be Continuous if you were measuring time of survival in Months, but you could have them set, for example, as Month on a treatment, then they would be Ordinal.

Appropriate charts: Bar charts
Appropriate descriptive statistics: Frequency table, Median, 1st and 3rd Quartile

**Relationship between 2 continuous variables:**

Appropriate charts: Scatterplot
Appropriate descriptive statistics: Correlation
Relationship between 2 ordinal or nominal variables:

Appropriate charts: Grouped Barcharts, Side by Side Barcharts
Appropriate descriptive statistics: Crosstabs

Relationship between 1 ordinal or nominal variable and 1 continuous variable:

Appropriate charts: Side by Side Boxplots, Stacked or Side by Side Histograms.
Appropriate descriptive statistics:
Grouped means and standard deviations.

Examples of Charts and Graphs:

Barchart
Crosstabs Table:

<table>
<thead>
<tr>
<th>Age</th>
<th>Unlisted phone number</th>
<th>No</th>
<th>Yes</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% within column</td>
<td>24%</td>
<td>49%</td>
<td>29%</td>
</tr>
<tr>
<td>18-34</td>
<td>n</td>
<td>185</td>
<td>90</td>
<td>275</td>
</tr>
<tr>
<td>35-44</td>
<td>% within column</td>
<td>20%</td>
<td>26%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>153</td>
<td>48</td>
<td>201</td>
</tr>
<tr>
<td>45-54</td>
<td>% within column</td>
<td>17%</td>
<td>10%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>133</td>
<td>19</td>
<td>152</td>
</tr>
<tr>
<td>55-64</td>
<td>% within column</td>
<td>17%</td>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>130</td>
<td>21</td>
<td>151</td>
</tr>
<tr>
<td>65+</td>
<td>% within column</td>
<td>23%</td>
<td>3%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>178</td>
<td>6</td>
<td>184</td>
</tr>
<tr>
<td>NET</td>
<td>% within column</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>779</td>
<td>184</td>
<td>963</td>
</tr>
</tbody>
</table>

Box Plots

![Box Plots](image)
Side by Side Box Plots

Histogram
Side by Side Histograms

distribution of height of 2 durum wheat varieties

Stacked Barcharts
All graphs taken from: https://www.r-graph-gallery.com/all-graphs.html
Cross tabs example from https://www.displayr.com/what-is-a-crosstab/