

Graphs

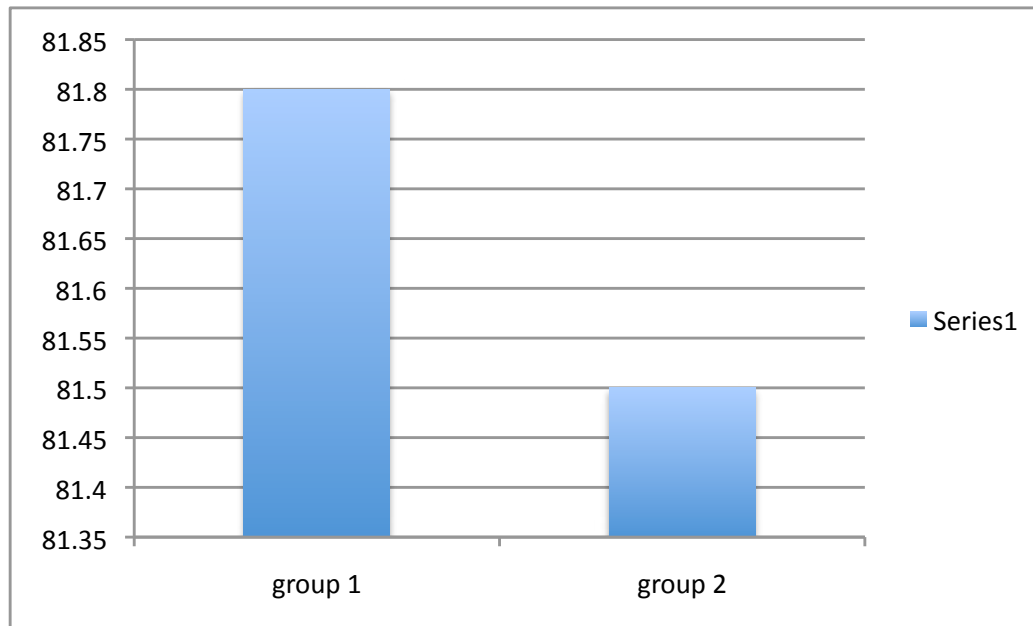
Types of Graphs

Histograms

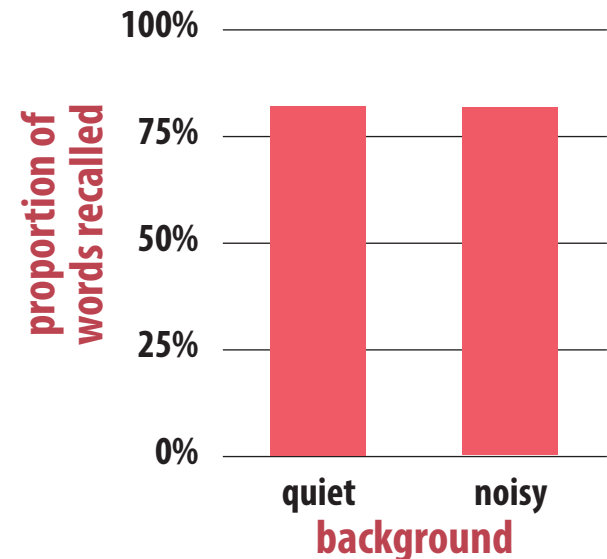
The purpose of a graph

The purpose of a graph is to make results clear at a glance. Avoiding common pitfalls, such as those done by default in Microsoft Excel (shown below on the left), can help achieve this goal.

| Problem with default Excel graph | Consequence |
|---------------------------------------|-------------------------------------|
| no x-axis label; untitled categories | not known what is being compared |
| no y-axis label or title | not known what is being measured |
| y-axis doesn't have appropriate range | misleading comparison |
| inconsistent decimal places | distracting, unprofessional |
| "Series1" | confusing because it is meaningless |



Effect of Noise on Recall

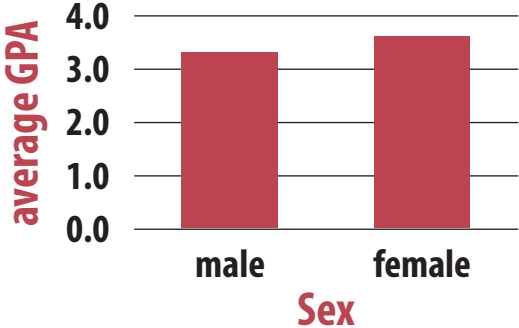
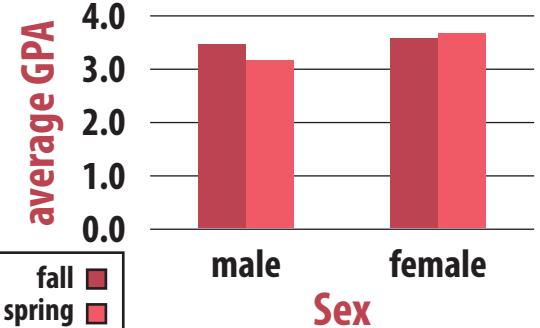
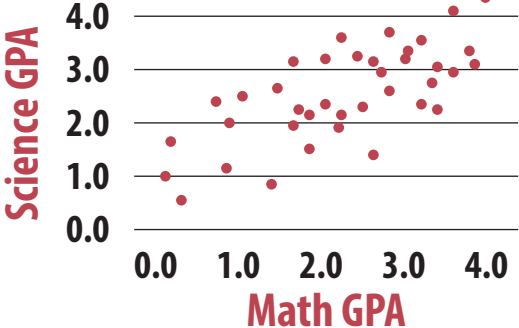
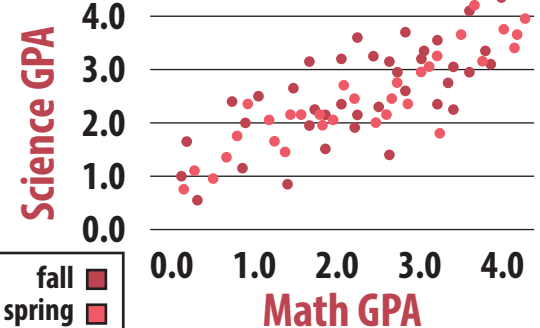
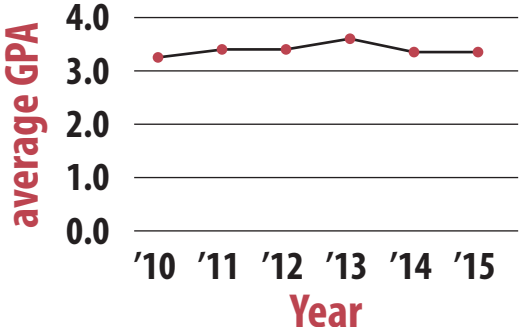
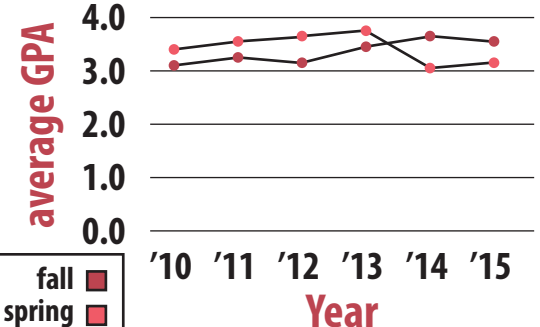


Types of graphs

| Graph | Description | Purpose | When to use |
|-------------------------|--|---|--|
| Bar Graph | one bar for each category | to compare averages, totals, or frequencies of categories | The independent variable is categorical, such as car brand. |
| Scatter Plot | one dot for each data point | to show the correlation between two numerical variables | Each participant provides two numerical values, such as height and weight. |
| Time Series Plot | one dot for each point in time, connected with line segments | to show a trend over time | The independent variable is time on a timeline, such as year, and the dependent variable is numerical, such as rainfall. |

A **pareto chart** is a bar graph in which the bars are presented in order from highest to lowest.

Graph Examples

| Graph | One independent variable | Two independent variables |
|---------------------------|--|--|
| <h2>Bar Graph</h2> |  <p>average GPA</p> <p>male female</p> <p>Sex</p> |  <p>average GPA</p> <p>male female</p> <p>Sex</p> <p>fall spring</p> |
| <h2>Scatter Plot</h2> |  <p>Science GPA</p> <p>Math GPA</p> |  <p>Science GPA</p> <p>Math GPA</p> <p>fall spring</p> |
| <h2>Time Series Plot</h2> |  <p>average GPA</p> <p>'10 '11 '12 '13 '14 '15</p> <p>Year</p> |  <p>average GPA</p> <p>'10 '11 '12 '13 '14 '15</p> <p>Year</p> <p>fall spring</p> |

Histograms and Distributions

A discrete **distribution** states each possible category and how common it is.

A **histogram** is a type of bar graph in which the x-axis is a numerical scale and the y-axis shows the **frequency** (how many) or **relative frequency** (percentage) of data in each category.

| Distribution | Description | Histogram Example | | | | | | | | | | | | | | |
|-----------------------------|--|--|--------------------|-------|-------|--------|-------|-------|--------------|------------|-------|-------|-------|-------|-------|-------|
| Uniform | Each category is equally common. | <p>6-Sided Die Rolls</p> <p>% of rolls</p> <table border="1"> <tr><th>Roll</th><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr> <tr><th>% of rolls</th><td>16.67</td><td>16.67</td><td>16.67</td><td>16.67</td><td>16.67</td><td>16.67</td></tr> </table> | Roll | 1 | 2 | 3 | 4 | 5 | 6 | % of rolls | 16.67 | 16.67 | 16.67 | 16.67 | 16.67 | 16.67 |
| Roll | 1 | 2 | 3 | 4 | 5 | 6 | | | | | | | | | | |
| % of rolls | 16.67 | 16.67 | 16.67 | 16.67 | 16.67 | 16.67 | | | | | | | | | | |
| Bimodal | The two most common categories are not next to each other. | <p>Starbucks Drink Prices</p> <p>% sold today</p> <table border="1"> <tr><th>Price (\$)</th><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr> <tr><th>% sold today</th><td>10</td><td>35</td><td>20</td><td>25</td><td>10</td></tr> </table> | Price (\$) | 1 | 2 | 3 | 4 | 5 | % sold today | 10 | 35 | 20 | 25 | 10 | | |
| Price (\$) | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | | |
| % sold today | 10 | 35 | 20 | 25 | 10 | | | | | | | | | | | |
| Approximately Normal | The distribution is approximately mound-shaped and symmetrical. | <p>Scores on Physics Test</p> <p># of tests</p> <table border="1"> <tr><th>Score</th><td>40-50</td><td>50-60</td><td>60-70</td><td>70-80</td><td>80-90</td><td>90-100</td></tr> <tr><th># of tests</th><td>5</td><td>10</td><td>15</td><td>18</td><td>12</td><td>5</td></tr> </table> | Score | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 | # of tests | 5 | 10 | 15 | 18 | 12 | 5 |
| Score | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 | | | | | | | | | | |
| # of tests | 5 | 10 | 15 | 18 | 12 | 5 | | | | | | | | | | |
| Skewed | The distribution has one or a few data far out on one specific side. | <p>Home Prices (\$100,000s)</p> <p># sold</p> <table border="1"> <tr><th>Price (\$100,000s)</th><td>0-5</td><td>5-10</td><td>10-15</td><td>15-20</td><td>20-25</td><td>25-30</td></tr> <tr><th># sold</th><td>10</td><td>20</td><td>12</td><td>5</td><td>2</td><td>1</td></tr> </table> | Price (\$100,000s) | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | # sold | 10 | 20 | 12 | 5 | 2 | 1 |
| Price (\$100,000s) | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | | | | | | | | | | |
| # sold | 10 | 20 | 12 | 5 | 2 | 1 | | | | | | | | | | |