From the following set of graphics choices, choose the graphic that would best be used to demonstrate the results of the study

Table 1: Graphics Choices to be used to answer questions 1 through 5

| A) Pie Chart | B) Bar Chart | C) Multiple Bar Chart | D) Pareto Chart |

Use the information given in the following setting to answer questions 1 through 5:

A sociologist has surveyed 345 people and recorded their gender, race, marital status, and job type. Table 2 identifies the variable and the possible “values” that the variable can have.

Table 2: Variables and their possible values:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Race</th>
<th>Marital Status</th>
<th>Job Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>White</td>
<td>Married</td>
<td>Teacher</td>
</tr>
<tr>
<td>Female</td>
<td>Black</td>
<td>Widowed</td>
<td>Counselor</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>Divorced</td>
<td>Electrician</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>Never Married</td>
<td>Plumber</td>
</tr>
<tr>
<td></td>
<td>Native American</td>
<td>Single</td>
<td>Carpenter</td>
</tr>
</tbody>
</table>

1) From the graphics choices given which might be used to illustrate the distribution of marital status?

Since the variable “marital status” is qualitative and there are only 5 categories you can use a pie chart to illustrate this distribution. Another choice here would be a bar chart.

2) From the choices given, which might be used to compare the distribution of race split by gender?

When the goal is to compare two or more populations (or sets of sample results) on a categorical variable use a multiple bar chart.

3) From the choices given, which might be used to compare the distribution of job type split by race?

Use a multiple bar chart because the objective is to compare results on a qualitative variable.

4) From the choices given, which might be used to illustrate the distribution of job type?

Since there are more than 6 classes of job type a pie chart would become too cluttered. Hence, use a bar chart.
5) If the investigator wanted to illustrate the job type variable from highest proportion in a job type to lowest proportion in a job type, which of the graphics choices would she use?

When the goal is to look at an ordered arrangement of the categories the Pareto chart is used. Recall that the Pareto chart is just a bar chart with the categories arranged so that the category with the highest frequency (or percentage) is first followed by the category with the next to highest frequency ... etc.

Use the information in the following setting to answer questions 6 through 8

A fish biologist working at a nearby fish hatchery has been following the growth of the season's new fry. (Probably with a time series graph 😊). One day she randomly selects 10 fry and records their lengths (cm). The data follow in Table 2.

Table 2: Lengths (cm) of 10 randomly selected fry

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9.5</td>
<td>10.3</td>
<td>10.7</td>
<td>10.9</td>
<td>11.4</td>
<td>11.5</td>
<td>11.9</td>
<td>11.9</td>
<td>12.2</td>
<td>12.5</td>
</tr>
</tbody>
</table>

6) Compute the average length of fry in this sample

\[ \bar{x} = \frac{\sum x_i}{n} = \frac{1}{10}(112.8) = 11.28 \approx 11.3 \]

7) Compute the median length of fry in this sample

1) First find the rank of the median \( Rank = \frac{(n + 1)}{2} = \frac{11}{2} = 5.5 \)

2) This tells us to go find the midpoint between the 5th and 6th data points. This value is the median.

\[ Med = \frac{(11.4 + 11.5)}{2} = 11.45 \]

8) Change the last observation from 12.5 to 125 and repeat steps 5 and 6.

\[ \bar{x} = \frac{\sum x_i}{n} = \frac{1}{10}(225.3) = 22.53 \approx 22.5 \]

Note that the mean changes quite a bit. In fact it is larger than 9 out of the 10 observations. The median, however, remains 11.45. This is an example of a robust statistic. A robust statistic is one whose value doesn't change too much even if there are outliers present in the data set. We say that the mean is not robust to outliers but the median is robust to outliers.