Things to consider when choosing an appropriate model for a given data set:
There are no exact "rules". Someone may choose one type of model, and provide good reasons for it, while someone else chooses a different model, and presents good reasons as well. Therefore, you should try to find a model that is reasonable to you, and provide your reasons, rather than trying to find "the correct model". There usually is not ONE correct model.

Draw a scatterplot.

(1) **Consider the shape:**
- Does the scatterplot look essentially **straight**? If Yes: → LINEAR MODEL
  If No:
  - Is there a **change in concavity**? (i.e. is there an inflection point?)
    If Yes: Only the cubic and logistic model can change concavity.
    If No: Your choices are probably exponential or quadratic.
  If there are many changes in concavity, re-consider linear. OUR models can handle only ONE change in concavity.

(2) **Consider the overall fit** of your first choice model to the data. Let's say, you decide to fit an exponential model. You then find it and graph it with the data. If the model you find does not fit the data well, you may want to try a quadratic (or other) model, although it was not your first choice. If your second choice imitates the data a lot better, this better fit may outweigh your reasons to choose the other model, such as end-behavior. (Same with logistic and cubic).

(3) **Consider the "end-behavior".**
   Does the graph seem to **level off**? Where? (right/ left/ both sides ?)
   Most-often, this will help you choose between the two options left.

Sometimes, it is not clear from the scatterplot whether the data levels off or not, but the **context indicates** that the data SHOULD (or should not) level off. For example, a graph of the world population should level off at 0 at the left, if the input is time, because the further back you go in time, the fewer people were on the planet...

(4) Always use the general rule of thumb: **Simpler is Better**.
   For example, if a linear model does not fit really well, and none of the others fits a whole lot better, go with the simplest one! Keep in mind that a more complicated model will always fit better because it can do more things. If the improvement in the fit to the data is not significant, go with the simpler model.

(5) Finally: We will **only use the logictis or the cubic** model if the data (scatterplot) shows an **inflection point** (a change in concavity).
For each of the scatterplots that follow: Use the rules from above to decide which model to choose for the plotted data. There is no context given. So you can ONLY use the data and your best judgement. Indicate which of the rules you used! Back up your choice as well as you possible can! This will require some writing - and a careful reading of the rules on page 1 above.
For each of the following three data sets:
(a) Discuss briefly which of our five possible models is or is not appropriate.
   Indicate which of the rules for selecting a model you used.
   (Formulate your reasons!)
(b) Find and report the model you believe is best choice. Why is it the best choice?
(c) Use your model from (b) to estimate the data point for the year 2000.
(d) Do you believe that your model is appropriate for this or other extrapolations?
   (The answer may not be the same for all three. Why not?)

(1) The table shows yard waste as a percentage of all waste generated in the US:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>yard waste (%)</td>
<td>18.2</td>
<td>18.2</td>
<td>17.1</td>
<td>14.1</td>
<td>13.4</td>
</tr>
</tbody>
</table>

(2) The table shows how many out of 1000 15-19 year old women gave birth in the years 1991-1997.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># births</td>
<td>16.3</td>
<td>15.9</td>
<td>15.5</td>
<td>15.2</td>
<td>14.8</td>
<td>14.7</td>
<td>14.6</td>
</tr>
</tbody>
</table>

(3) The table gives the death rates (deaths per 100,000 males in US) due to lung cancer from 1930 through 1990:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>deaths per 100,000</td>
<td>5</td>
<td>11</td>
<td>21</td>
<td>39</td>
<td>59</td>
<td>66</td>
<td>67</td>
</tr>
</tbody>
</table>