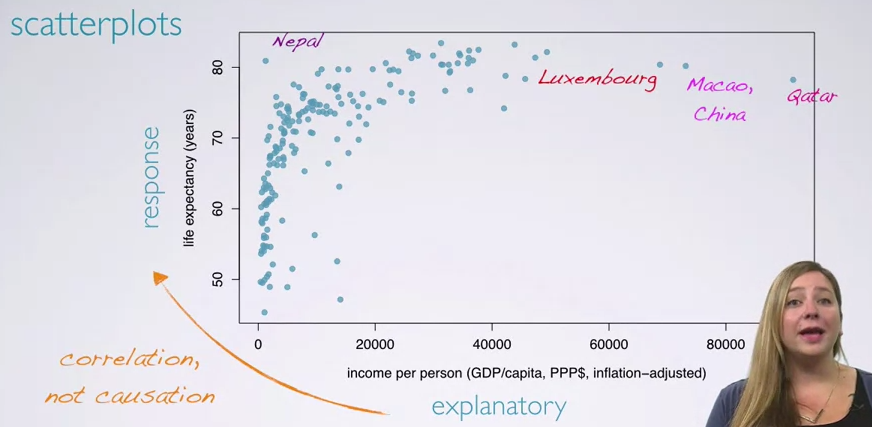
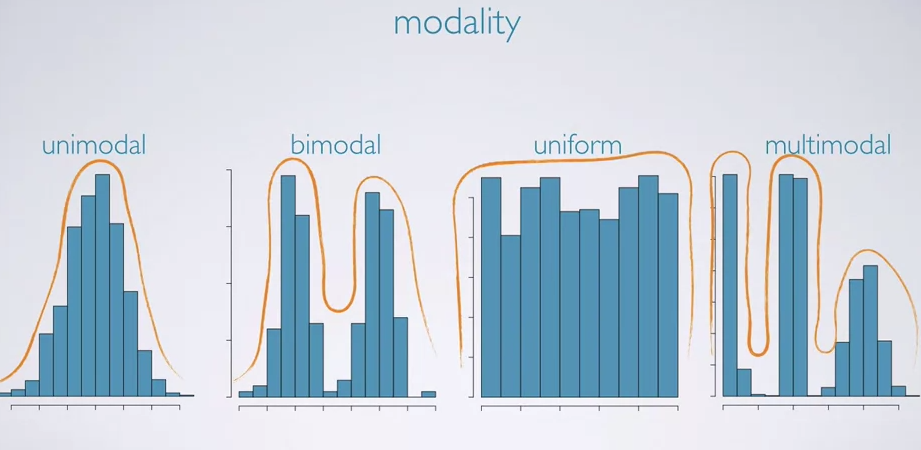
Coursera Stats Video Lectures – Visiualizing Numerical Data

Week 1, Video 6 – Video Source – Gapminder.com

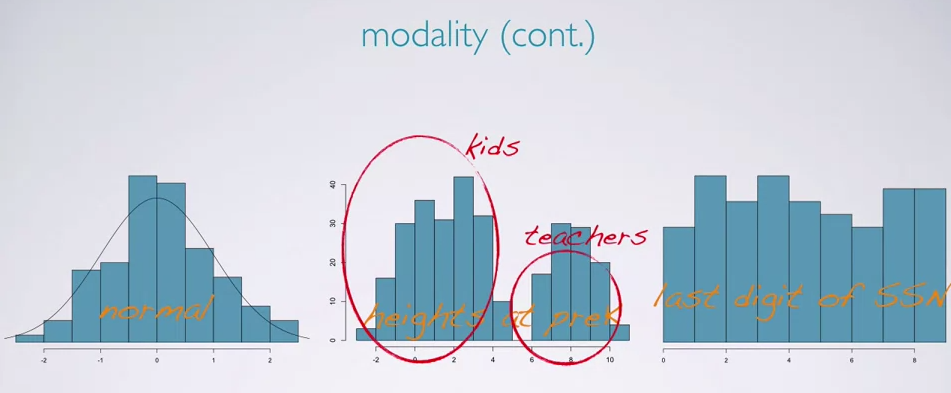
1. Scatterplot for paired data
   1. explanatory is the x variable, y is the response variable
   2. no evidence of causation at this stage, only correlation



1. evaluating variable relationship
   1. direction of line of best fit
      1. increasing, decreasing
   2. shape of line
      1. linear, exponential
   3. strength of relationship
      1. weak (lots of scatter), strong (very little scatter)
   4. outlyers
      1. observations far away from the rest of the observations and the line of best fit
      2. outlyers: don’t immediately exclude them
2. other methods of visiualizing data
   1. histogram
      1. useful for numerical data
      2. provides view of data density
      3. useful for describing shape of distribution
   2. Skewedness
      1. left skewed: long tail to the left; negative end
      2. right skewed: long tail to the right, positive end
      3. symmetric: no skew to left or right
   3. Modality

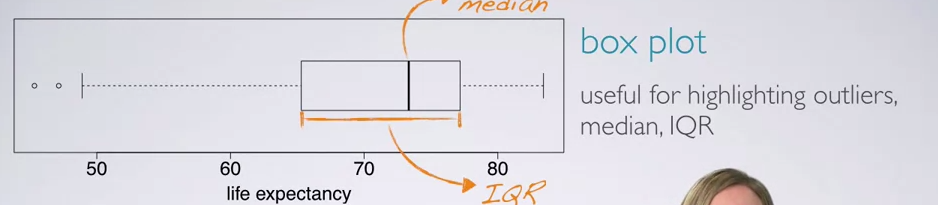


* 1. most common modality in stats is unimodal (bell curve) and bimodal.



* 1. uniform distributions imply the variable is just as likely to have a low value as a high value

1. dotplots
   1. used when individual values are of interest, not good for large samples
2. boxplot
   1. useful for highlighting outlyers, median, and the interquantile range



1. Intensity map: showing spatial distributions
   1. such as showing income differences on a map