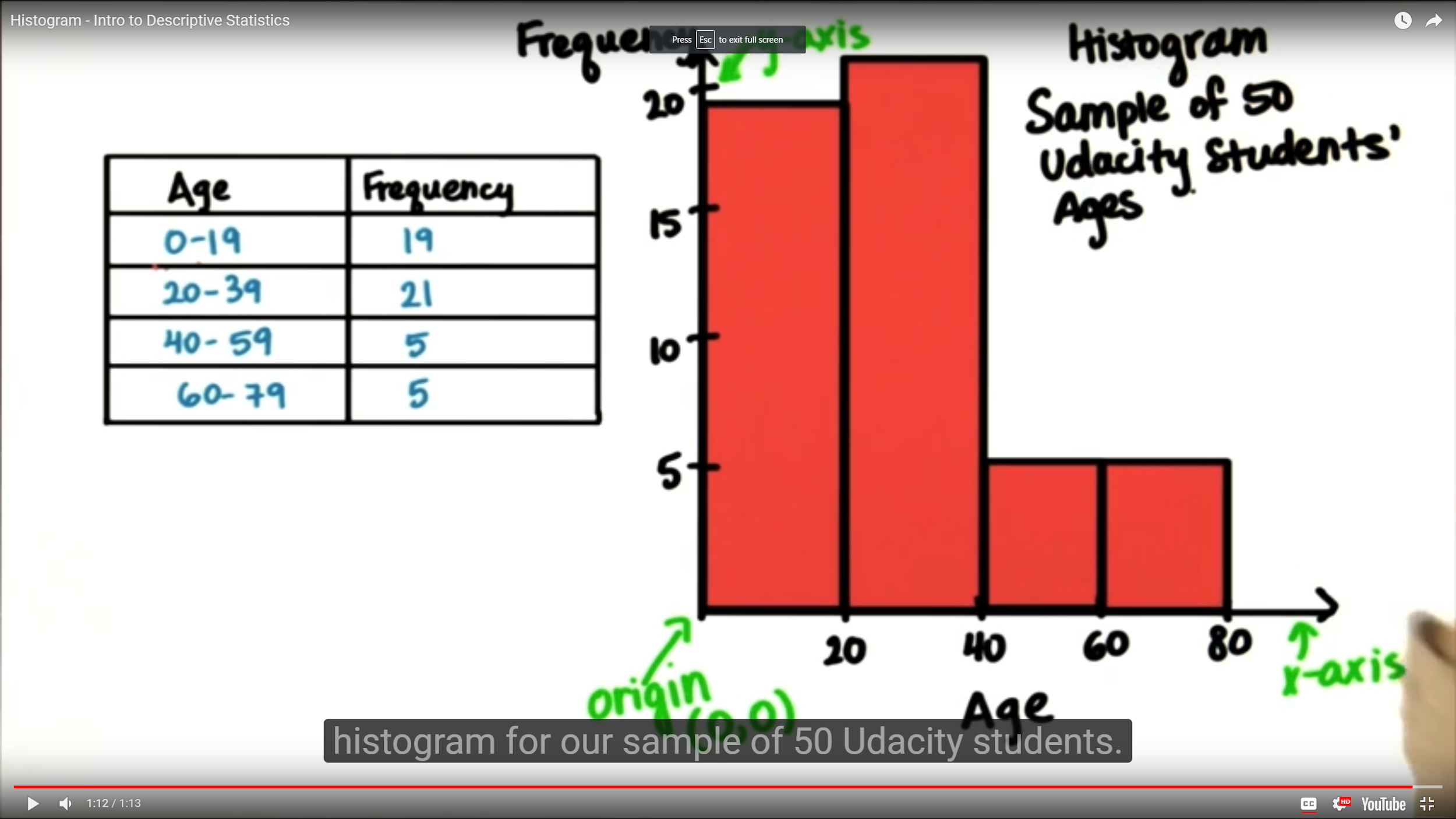
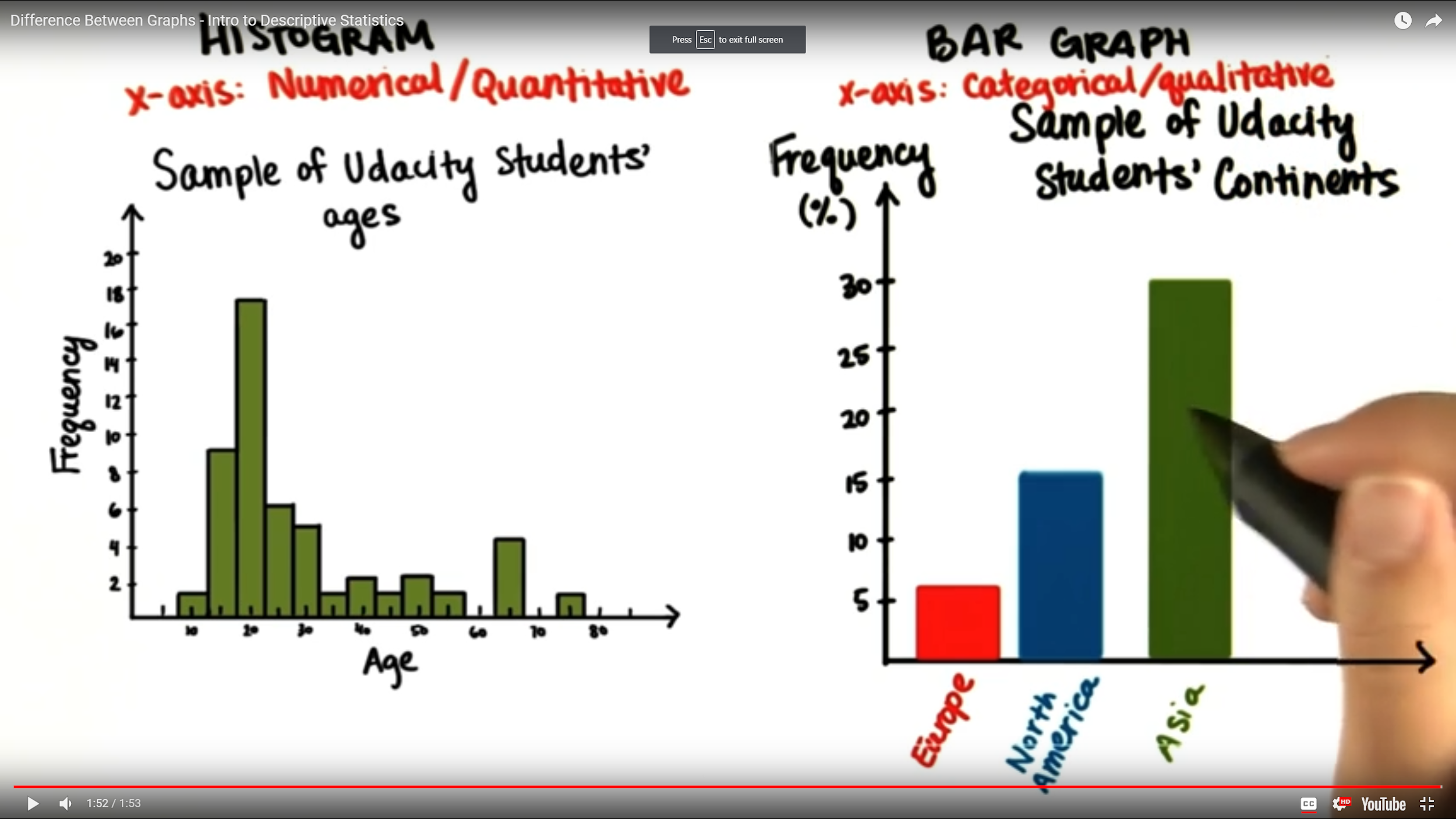
BS – Important Concepts Visualizing Data – Lessons 6, 7, 8  
  
1. Frequency – frequency table:  
-> good to visualize data;   
-> easier to understand;   
-> quickly analyze data;   
-> see patterns and make decisions  
-> y-axis  
-> good for calculating the total number in the sample/population (n);  
-> we have exact counts, so we can always create the histogram, add the frequencies in each bin   
2. Relative frequency – how they relate to each other   
3. Proportion:  
-> should add to 1   
-> greater than 0 and less than 1, or may be equal to either of these values  
4. Percentage:   
-> proportion \* 100  
-> greater than 0 and less than 100, or may be equal to either of these values  
5. Bin/interval/bucket size:   
-> group data;  
-> it depends on how much detail you want;   
-> interval in which you're counting the frequencies;  
->sometime sacrifice detail for convenience  
-> bin size too small – too much detail  
-> bin size too big – not enough detail  
-> when we visualize data with histograms, as we make the bin size bigger, the frequency gets larger. As we make the bin size bigger, more values will fall inside that bin.   
  
  
 Difference between Graphs: Histogram – Bar Graph  
6. Histogram:

-> shape is important

-> var is numerical, quantitative;

-> good for visualizing the shape of the distribution;

-> can always be created if you have the frequency table;

-> you don't always know the exact frequencies   
7. Bar graph:

-> distinct categories

-> order doesn't matter

-> shape not important – arbitrary

-> categorical/qualitative variables

-> overall percentages  
8. Normally distributed data:   
-> has one peak around the mode;   
-> the shape is roughly symmetrical;   
-> most scores fall in the middle of the distribution  
-> the mean, median, and mode coincide  
9. Skewed distribution:   
–> positive distribution: scores with the lowest frequencies are on the right side of the distribution;  
-> negative distribution: a left-skewed distribution has a long left tail. Left-skewed distributions are also called negatively-skewed distributions. That’s because there is a long tail in the negative direction on the number line. The mean is also to the left of the peak

Link:

Interactive histogram software - Interactive Histogram Applet   
http://www.shodor.org/interactivate/activities/Histogram/