Summary of main insights of Data Visualization

- Categorical data:
  There are too many categories for the some attributes which can be avoided if we club some similar matching categories into a big category.
  For example, in NATIVE- COUNTRY as the count of United-States is about 90%, the other countries, although there have no significant impact. If we club those countries by continents, they would have much better ratio against United-States.

- Continuous data:
  The histograms of continuous attributes with different bin width give us insights on how the data is skewed:
  - Age : Partially left skewed
  - FNLWGT: Left skewed
  - Education-num: Right skewed
  - Capital-Gain: Left skewed
  - Capital-loss: Left skewed
  - Hours-per-week: Gaussian Distribution

  Also the attributes 'Fnlwgt' and 'Education-num' are better removed from the dataset as 'Education-num' is just the duplicate of the 'Education' categorical attribute.