CSE 5243 – Introduction to Data Mining

Rajarshi Biswas, Sayam Ganguly

{biswas.91, ganguly.32}@osu.edu

The Ohio State University

# 1 Exploratory analysis of the Income Dataset

The dataset is an extraction from the 1994 Census database. From the income point of view the data is largely classified between more than 50K and less than 50K.

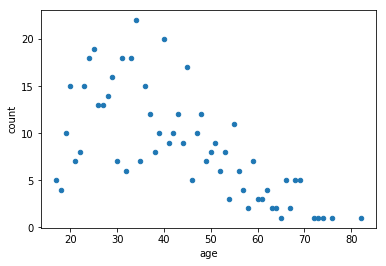
## 1.1 Overview of the income Dataset

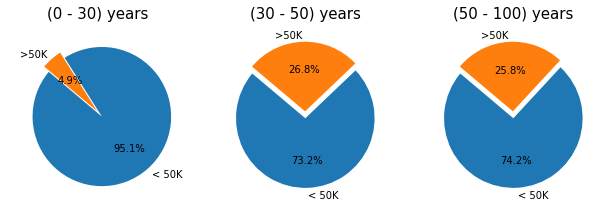
The income dataset has total 520 entries. And the following columns. 28 rows have missing values.

* age: continuous.
* workclass
* fnlwgt: continuous. Meaning is ambiguous.
* Education
* education-num: continuous.
* marital-status
* occupation
* relationship
* race
* gender
* capital-gain: continuous
* capital-loss: continuous.
* hours-per-week: continuous.
* native-country
* class: >50K, <=50K.

## 1.2 Analysis

### 1.2.1 Age





### 1.2.2 Workclass

