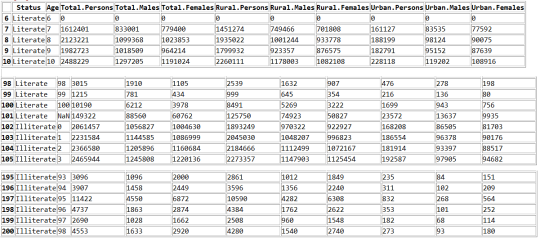
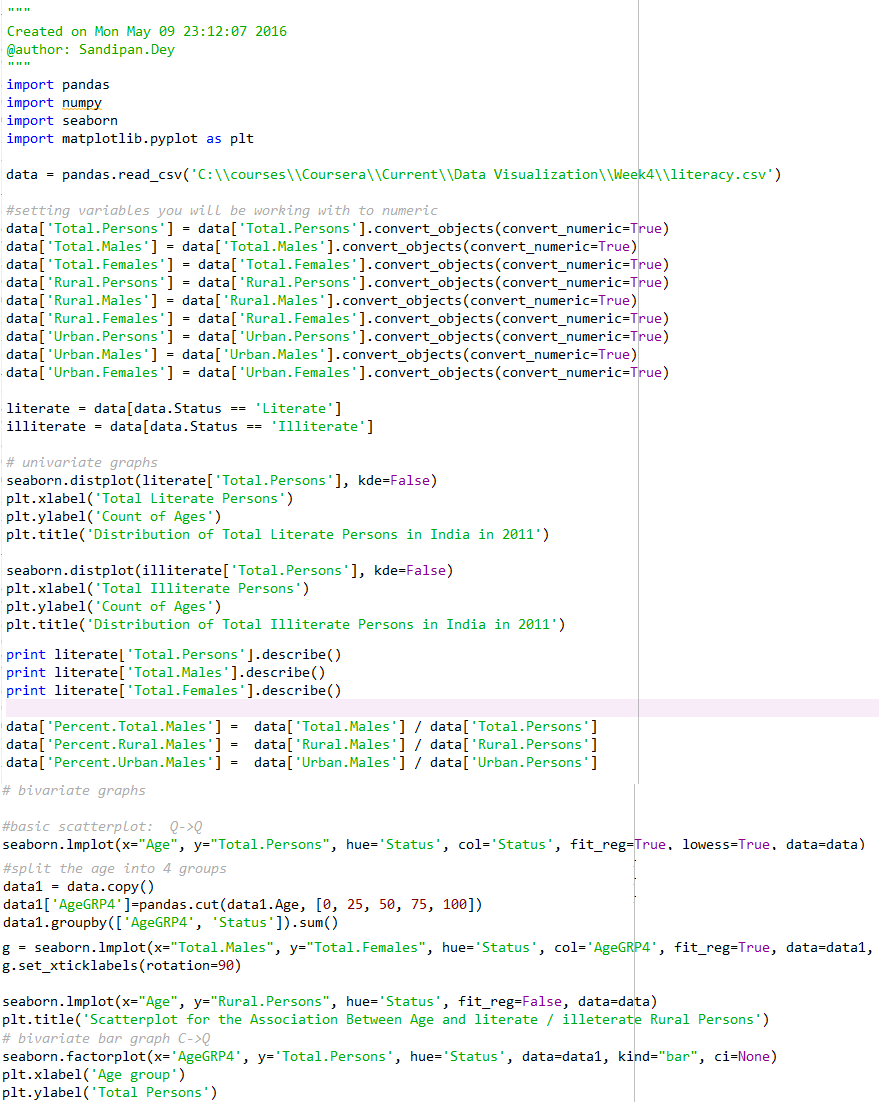
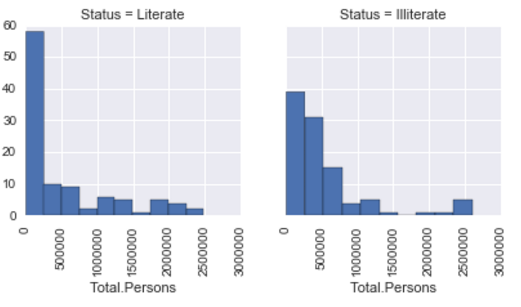
Literacy Dataset from Census (2011) in India: Exploratory Analysis with graph plots

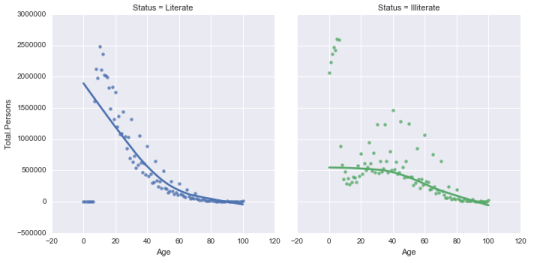
The dataset is available as the **Age** data (Census India 2011) and can be found and downloaded from the following url: <http://www.censusindia.gov.in/2011census/population_enumeration.aspx> .   
  
A subset of the data was taken and analyzed to understand the association between literacy and the other variables such as age, gender, urban / rural aggregated across the country level (not across state levels).

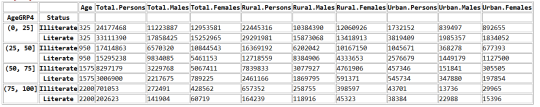
For each age in between [0,..,100], this subset contains couple of rows: one corresponding to the total number literate males/females at rural/urban areas in India and the second one corresponding to the total number of illiterate males/females.

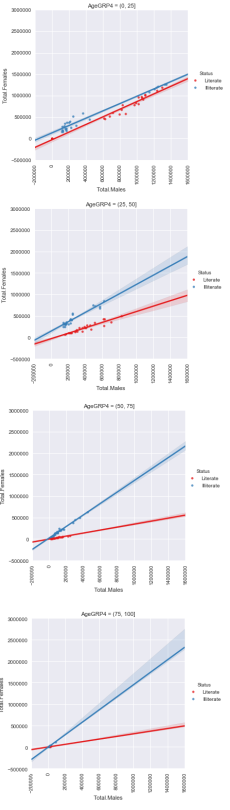
The dataset was first processed a little bit (by renaming variables) to obtain the following subset:



**Python Code**  
  
**Output**  
  
As can be seen from the following histogram plots, for around 60 different ages, the total number of literate persons is in the interval [0, 250000], whereas for around 35 different ages the total number of illiterate persons is in the same interval.    
  


The following shows the scatter plots (and the corresponding *lowess regression*fitted lines) show the associations in between the age and the total number of persons separately for the literate and the illiterate persons. As can be seen, the total number of literate persons are much higher for low age values and drops rapidly as the age grows.  
  


Binning the variable *Age*into 4 age groups [0,25], [25,50], [50,75] and [75,100] and grouping by the Age Group and the Status variable, we get the following table.  


The following graphs show the scatterplots of the variables total females and total males for different age groups.  


The following figures show the association between age and total number of rural and urban persons for literate and illiterate persons.  


The following figures show the comparative barplots for the literate / illeterate persons for different age groups for rural and urban persons.  
