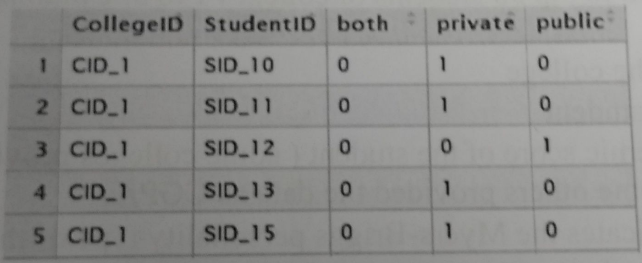
Questions:

1. Read the files relating to top 50 students from all the 5 colleges
2. Observe the data distribution of each column in all the 5 dataframes. For each college,
   a. Report if the distribution of numeric variables is normal. ***Hint: Can we identify this by observing the summary or some kinds of plot reveal it.***

   b. Also report if any one category in categorical attribute has dominance over other categories. ***Hint: Can this be answered by observing the counts of each level***

   c. Find attributes that have missing values

   d. Report how many missing values are present in each file 52 → All (9,6,19,9,9)
3. In each of the files fill missing values
4. Combine all those data frames into a single, consolidated data frame, and name it as *"consolidated_data"*
5. Read the Placements.csv and observe the data, and the format in which it is given
6. Transform the data into this format using reshape Try



	CollegeID	StudentID	both	private	public
1	CID_1	SID_10	0	1	0
2	CID_1	SID_11	0	1	0
3	CID_1	SID_12	0	0	1
4	CID_1	SID_13	0	1	0
5	CID_1	SID_15	0	1	0

7. Merge this data with the above consolidated_data, properly
8. Derive an attribute named "*isPlaced*" that contains a zero if the student is not placed, and that contains 1 if he is either placed in public/private sector companies, or in both. Achieve this using apply and if-else functions
9. How many students from each college, are placed in both private and public-sector companies? Irrespective of the type, how many students are placed in each college?
10. Find how the mean overall_score is faring across various extra-curricular activities. Which plot would be appropriate