DATA EXPLAIN

In this project I will test the overconfidence hypothesis using a different set of data. The variables included in the dataset are:

- *year*: Year of the observation
- *firmid*: A firm specific identifier
- *I*: Investment (%), defined as firm capital expenditures as a percentage of capital at the beginning of the year
- *CF*: Cash flow, defined as earnings before extraordinary items plus depreciation as a percentage of capital at the beginning of the year
- *MV*: Market value of assets over book value of assets at the beginning of the year (%)
- *S*: Percentage of total shares owned by CEO or CEO's family (%)
- *Size*: Logarithm of assets (£) at the beginning of the year
- *CG*: Number of outside directors who are CEOs in other companies
- *TC*: Press based overconfidence measure, the number of articles mentioning CEO as "confident" or "optimistic"
- *TD*: A dummy variable equal to 1 if the majority of articles portrayed confidence and 0 otherwise