

## ST300 Project: data description

### Variables in the dataset

Variable	Data type	Description
Region	factor	Levels are US: USA, EUR: Europe, EMG: Emerging markets
Industry	factor	List of sectors
Institutional_holding	numeric	Percentage of shares held by institutions
Number_of_firms	Numeric/integer	Number of firms in each sector
EPS_Growth	numeric	Earnings-per-share growth in %
PBV	Numeric	Price-Book value
Beta	numeric	Estimated from CAPM
CEO_holding	numeric	Percentage of shares held by CEO
Cost_of_Equity	numeric	In percentages
ROE	numeric	Return-on-equity (%)
PE	numeric	Price-to-earnings ratio
PS	numeric	Price-sales ratio

The data file contains a header.

At the start you need to do some data pre-processing. There are some variables that are numeric but regarded by R as factors. To guide you, see the Get Started.Rmd in Moodle.

### Learning more about the variables

You do not need in-depth knowledge of the financial variables. Some of the variables are self-explanatory. A video explaining the PE ratio is given by Prof Damodaran:

<https://youtu.be/42iyR6Gegiw>

He also has videos explaining other financial ratios.