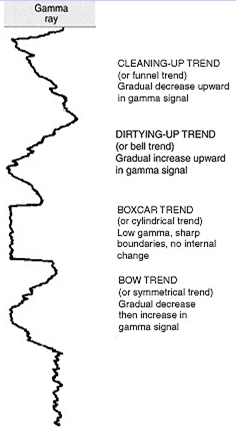
## **Gamma Log Facies Type Prediction**

Data Distribution:

* Train data: 4.4 million rows
* Test data: 2.2 million rows

Columns:

* **well\_id**:
  + Each well\_id identifies a unique well
  + The train file contains data for 4000 wells numbered from 0 to 3999
  + The test file contains data for 2000 wells numbered from 5000 to 6999
* **row\_id**:
  + Corresponding to each well there are exactly 1100 rows numbered from 0 to 1099
  + The numbering of the rows corresponds to the depth profile of the Gamma Ray log
* **GR**:
  + Gamma-Ray Log values whose plots need to be interpreted
  + Each portion of the GR value plot has a specific geometry based on which they can be assigned a label



Source of image: Emery (1996): Character of Log Response; [www.sepmstrata.org/page.aspx?pageid=168](http://www.sepmstrata.org/page.aspx?pageid=168)

* **label**:
  + This is the variable that needs to be predicted
  + The values **categorical and nominal.**
  + There are 5 classes, each corresponding to specific trend in the GR value plot:
    - 0 (None),
    - 1 (Symmetrical),
    - 2 (Cylindrical),
    - 3 (Funnel),
    - 4 (Bell)

Source of image: Kendall, C.G.St.C. (2003): Character of Log Response; www.sepmstrata.org/page.aspx?pageid=168.


Source of image: Kendall, C.G.St.C. (2003): Character of Log Response; [www.sepmstrata.org/page.aspx?pageid=168](http://www.sepmstrata.org/page.aspx?pageid=168).