# Data explorations, event detection, and introduction to statistics

**1- Load the library welly in the conda environment DrillingAnalytics and launch Jupyter Labs**

**2- Load welly library in notebook**

**3- Load well log in welly object: 8267\_a0801\_1996\_comp.las**

**4- Plot well log data**

**5- How many GR values are in the log and how many are missing?**

**6- What is the average of GR values?**

**7- What is the average of GR values in a the sand?**

**8- Determine which curves have no data, which have gaps, flatlined and if GR is below 25**

**9- What is the proportion of NaN values in each curve?**

**10- Which curves have spikes?**

**11- Load the gamma ray curve on to a dataframe**

**12- Load all the curves on to a dataframe**

**13- How many gamma ray readings are there in the dataframe? How many readings should there be for the entire well? If there is a difference, why would that be?**

**14- Define a Curve class to hold all the curves in a well log**

**15- Instantiate the object with gamma ray data**

**16- Print the name of the well, the location, the curve names and data using attributes**

**17- Print the name of the well, the location, the curve names and data using methods**

**18- Print the name of the well, the location, the curve names and data from 30 to 3000 meters using methods**

**19- Print the name of the well, the location, the curve names and data from 30 to 3000 meters using attributes**

**20- Load the las file using lasio and confirm the number of rows is correct: Data/LAS/8267\_a0801\_1996\_comp.las**

**21- Load a csv file on to a dataframe, making sure the name of the columns are correct: Swell-1A\_AsciiDrillData\_183.0-5006.csv**

**22- Determine if there are spikes in the data**

**23- Change all -999.25 readings to NaN**

**24- Determine how many NaN readings are there in each curve**

**25- Replace all NaN with the average of the before and after value. Compare the before and after**

**26- Replace all NaN with the average of the before and after value considering they could also be NaN values. Compare the before and after**

**27- Show the use of the rolling method in pandas**

**28- Filter all curves using rolling average**

**29- Load time LAS files for an entire run**

**30- Show the depth curve**

**31- Determine if the depth curve increases constantly**

**32- Graph the depth curve where it does not increment constantly**

**33- What are the RPM ranges?**

1) Graph the RPM distribution:

2) Delete the distribution around zero:

3) Graph with library sklearn

4) Compute normal values:

**34- Graph the normal distribution and show 95% confidence level ranges**

**35- Computer the statistical values for the normal values of RPM**

**36- Graph the normal distributions for the normal values of RPM using a a range of confidence level of 95%**

**37- Graph all the normal values for RPM in one graph**