Homework 4

when you don't create things, you become defined by your tastes rather than ability.
your tastes only narrow & exclude people. so create.

Jonathan Gillette (a.k.a. Why The Lucky Stiff)

You have become a great data analyst, and one of the major players in the Oil & Gas industry is requesting your services. They are starting to explore a new field, and they are sending you the data collected from the first well.

The reservoir engineer knows you are proficient in R, so he just wrote down the tasks you have to perform. Please follow them.

Pumpjack in West Texas
(source: wikimedia.org).

1. The well data is available for you to download here. Load the file into a convenient data structure.
2. The data you have received contains 7 variables. They are:
    - **X:** X-coordinate
    - **Y:** Y-coordinate
    - **Z:** Z-coordinate
    - **facies:** distinct rock types
    - **density:** rock density
    - **porosity:** rock porosity
    - **permeability:** rock permeability (measured in mD)

    To avoid problems in the future, make sure each column is in the appropriate data type.
3. Facies correspond to different zones in the reservoir. Identify how many samples you got from each zone.
4. As a first approach, to identify high pay zones, summarise all variables by facies, so that you can know what the average values are in each facies.
5. Another strategy is to find the samples where you have higher porosity values (more than 0.3). Identify those samples as well.
6. Meanwhile, you received more data that came up late. It was sent to you through the following link. This new data set has the following variables:
    - **X:** X-coordinate
    - **Y:** Y-coordinate
    - **Z:** Z-coordinate
    - **Vp:** P-wave velocity (km/s)
    - **Vs:** S-wave velocity (km/s)

      You were told the data acquisition process had some problems, and that it is possible you receive the same sample more than once. Take the necessary procedures to correct that.
7. With that issue eliminated, you are now able to gather the two data sets. Put them together according to the variables they have in common.

8. You have accomplished a great deed so far. Although you have gathered all available data, you can help the next step of the reservoir modelling by providing two more useful variables. The acoustic impedance is the product of density and seismic velocity, which varies among different rock layers. Add two columns to your current data set:
    - Ip: acoustic impedance considering the P-waves
    - Is: acoustic impedance considering the S-waves
9. Your job is finished, now you just have to deliver it. Save your data set in an appropriate file format.