1. In lecture I drew normal probability plots using the Savona.csv data set. This data set has been included along with the exercises in order for you to complete this question.

   (a) Draw the normal probability plot that I demonstrated in lecture.

   (b) Using the argument discussed in lecture, draw a lognormal probability plot of the data.

   (c) Sometimes it can be useful to compare two plots right next to each other. To do this, we can set graphing parameters in R using the function `par()`. If we type `par(mfrow=c(2,1))` Our graphics window will display 2 graphs, one on top of the other. Type the command `par(mfrow=c(2,1))`, and then type the commands to draw the normal and lognormal probability plots again to view them together.

   (d) Based on comparing the normal and lognormal plots right next to each other, would you choose to conduct the remainder of an analysis on the original scale, or on the log-scale?

2. Being able to identify the *population* that you are sampling from is key to knowing what inferences can and cannot be legitimately drawn from a data set. Below are some sampling scenarios; see if you can identify the population that we can draw inference to based on descriptions of samples that were taken.

   (a) You are creating a stage record for a river. You measure the river stage once a day, but on days where there is high precipitation or snow melt, the local bridge closes due to flooding so you cannot take an observation.

   (b) You are interested in contamination caused by a mining operation. What inferences can be drawn under the following sampling scenarios?

      i. You sample from the tailing pond only.

      ii. You sample from a lake near the tailing pond, and the tailing pond.

      iii. You sample from the lake only.

      iv. You sample from the lake, and groundwater sources near the tailing pond.

   (c) You are looking at a report submitted as part of a permit for a mine plan. The particular focus is the adequacy of the reservoirs submitted for the mine plan. The report includes 60 years of precipitation data, but the reservoirs are supposed to be designed to withstand a 1 in 200 year flooding event. Is the precipitation record being used for reservoir estimation a sufficient sample to draw inference to our population of interest?

   (d) A permittee is applying to expand a landfill site this year. Local community members are concerned that landfill leachate might affect their nearby water sources. What inference can you draw about the impact of leachate in community water sources under the following circumstances?

      i. Groundwater sampling occurred at the site in 2004 and 2005.

      ii. Groundwater sampling occurred between 2004 when the site opened, and the present within a 200m buffer zone around the landfill . The nearest residential well site is 150m from the landfill.

      iii. Same as above, except the nearest residential well is 500m away from the landfill.

3. Refer back to the Cadmium data set that was used in the third set of lectures and Exercise 3. We will be using it again here!
   In lecture we discussed that there are 3 main steps to creating good summary statistics. Following these steps, conduct a summary data analysis of the Cadmium data set. Justify any transformations of the data that you take. If you cannot complete the analysis due to assumption violations, discuss this.
   If you perform a data transformation, don't worry about transforming the summary data back to the original scale!