Lab Exercise 7

1)

Load **viperdb.csv** with Pandas.

Call simple Pandas functions to achieve the following tasks:

A)

Find the virus record with the largest outer capsid radius.

B)

Calculate the correlation between inner and outer capsid radius.

C)

Which viruses are larger? dsDNA or ssRNA?

D)

Assuming that viral capsids are roughly spherical, their surface area can be calculated by the formula $A = 4\pi R^2$ where R is the outer radius. Calculate this term for all virus records and add it to the DataFrame as a new column. How is it correlated with the **Outside SASA** field?

2)

A)

Download the PDF file **cancer_facts.pdf** and extract the table "Estimated Number of New Cases for Selected Cancers by State, US, 2014" into a CSV file. Use an online PDF-to-Excel tool such as: https://www.coolutils.com/online/PDF-to-CSV.

B)

Which state has the highest proportion of new Leukemia cases as a fraction of all cancer types?

C)

What is the percentage of new cancer cases not classified as one of the provided types per state?

D)

Find the two states that are most similar to each other with respect to the profile of new cancer types. Compare the fraction of cancer types, not the absolute values.