Lab Exercise 8

Throughout all analyses, make sure to name the axes and give your plots proper titles. Try to make your plots as aesthetic and clear as possible (use the style you find to be the most appealing).

1)

A)

In the previous Homework Exercise you worked with pfam domains in the human proteome.

Plot the distribution of domain lengths using the following chart types:

1) Histogram
2) CDF
3) Box plot

Plot the three graphs side by side and provide them with a shared title.

Which parts of the box plot you could have drawn yourself by merely looking at the CDF plot?

B)

Plot the length distribution of different HMM types in a single box plot (drawing a separate box for each type). Make sure to include in the labels and the total number of HMM occurrences of each type.

C)

Use a bar plot to show the average and standard deviation of domain length for each Pfam clan. Show only the 20 most common clans.

D)

Use a pie chart to show the number of Pfam domain occurrences falling into each of the 7 most common clans. Include an 8th label for “Others”.

E)

Use a scatter plot to show the distribution of E-values vs. the lengths of domain occurrences. Color the domains by their clan (use a different color for each of the 7 most common clans, and two more colors for other clans and domains without a clan).

F)

Use a line plot to show the E-value of occurrences of the **PDZ** domain in the **Q8NI35** protein as a function of the domain's position in the protein (given by the **alignment_start** attribute).