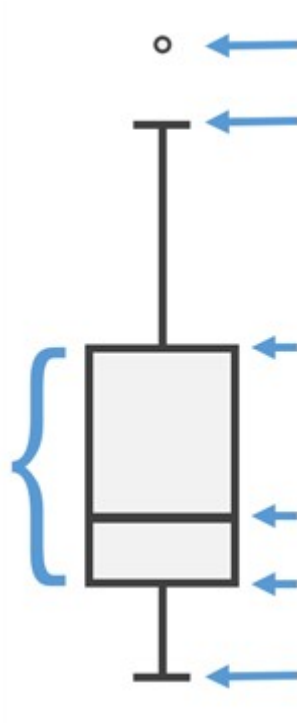


1. Egészítsd ki, mit jelölnek a boxplot különböző részei.  
(7p)



2. Írj egy függvényt, ami két int/float listát vár bemenetként és elkészíti ezek unióját, mint egy növekvően rendezett listát.

Pl.:

L1 = [2, 5, 0]

L2 = [3, 4, 5, 2]

merge(L1, L2) = [0, 2, 3, 4, 5]

(10p)

3. Készíts egy month\_dict szótárt, amelyben a kulcs a 'honap neve', 'honap szama' és 'napok szama'. A szótár értékei 12 elemű listák legyenek a hónapokkal. (4p)

Ezután készítsünk egy pandas dataframe-t a szótárból month\_df néven, aminek az oszlopai a szótár kulcsai. (3p)

Milyen függvényeket tanultunk amivel ugyanezt a dataframe-t excel-ből be tudtuk volna olvasni? (2p)

Válasszuk ki lekérdezéssel, melyik a legkevesebb napos hónap (csak a nevét). (2p)

Határozzuk meg átlagosan hány napos egy hónap. (2p)

Hány napos a leghosszabb nevű hónap? (2p)

4. Töltsd be a titanic datasetet sns-ből a tanult függvénnyel egy dataframe-be. (1p)

Milyen típusú a class illetve alone oszlop? (2p)

Hány különböző árban bocsátottak ki és összesen mennyi jegyet (fare)? (3p)

Ábrázoljuk a kor (age) függvényében a jegy árakat (fare), úgy, hogy azt hogy túlélők-e (alive), külön színű pontokkal jelöljük, továbbá egy 3x2-es grid-en jelenítsük meg, ahol osztály (class) illetve nem (sex) szerint szétbontjuk 6 részre az adathalmazt. (6p)

Mit állapíthatunk meg ez alapján a túlélőkről? (2p)

Ábrázoljuk megint a kor függvényében a jegy árakat és erre illesszünk lineáris regresszióval egy egyenest a konfidencia intervallum nélkül. Illesszünk ugyanerre másodfokú görbét. (4p)