

Data Science alapok - Segédanyag

Adattípusok:

1. Szeretnénk létrehozni egy számvektort, ami 50000 elemből fog állni. Mikor érdemes integer típust használni és mikor numeric típust?
2. Létrehozunk egy vektort 1 elemmel, ami csak egész számokat tartalmazhat. Milyen típusú legyen?
3. Mi a különbség a character típus és a faktor típus között?
4. Hozz létre egy character vektort, ami a következő szöveget tartalmazza: `single: ' double:"`
5. Faktorok esetén hogy tudjuk megkapni a kódolt értékeket (egész számok): `fac <- factor(c("Hello", "World"))`
6. Az alábbiak közül melyek használhatók igaz/hamis értékek reprezentálására? `true, TRUE, F, FALS, t, True`

R objektumok:

1. Mi a különbség a vektor és a lista között?
2. Mi a különbség a mátrix és a `data.frame` között?
3. Mi történik ha egy `data.frame` készítésekor az egyik oszlopvektor hosszabb mint egy másik oszlopvektor?

Típuskényszerítés:

1. Milyen típusúak lesznek az alábbi vektorok:

```
## [1] 1 2 3
```

```
## [1] 1 9
```

```
## [1] 0 1 5
```

```
## [1] "4" "3" "2"
```

```
## [1] 3+0i 4+0i 3+4i
```

```
## [1] "3" "char"
```

Explicit konverzió:

1. Számokat véletlenül faktorként olvastunk be. Hogyan tudjuk visszakapni a numeric értékeket? 2.a: Alakítsd a következő mátrixot `data.frame`-re: `mat <- matrix(1:6, nrow = 2)`. Hogyan tudnád megváltoztatni az oszlopneveit?

- 2.b: Ellenőrizd, hogy valóban numeric típusú!

Hiányzó értékek:

1. Mi `NA = NA` értéke és miért?
2. Hogyan tudjuk ellenőrizni, hogy egy érték hiányzó (`NA`)?
3. Milyen típusa lehet az `NA` értékeknek?
4. Van olyan művelet, amit `NA`-val elvégezve nem `NA` értéket kapunk? Miért van ez?
5. Töröld ki az `NA` értékeket a következő vektorokból:
 - `c(NA, 3, 5, 7, 11, NA, 3)`
 - `c(5, 10, 3, NA)`
 - `c(NA, NA, NA)`
 - `c(8, 9, NA)`
6. Imputáld (töltsd fel a hiányzó értékeket) a mediánnal a következő vektorban: `c(3, 9, 10, 5, 7, 8, 10, 2, 4, NA, 6, NA, 4)`
7. A következő mátrixnak:
 - Távolítsd el az összes olyan sorát, amiben hiányzó értékek találhatók
 - Válaszd ki az összes olyan sorát, amiben szerepelnek `NA` értékek

```
mat <- matrix(c(1, 2, NA, 4, 5, 6, NA, 7, 8, 9, NA, 10, 11, NA), nrow = 7)
```

NULL

1. Hozz létre egy listát 3 elemmel, majd töröld a másodikat.
2. Tartalmazhat-e egy vektor `NULL` értékeket? Miért?
3. Mi a különbség az `NA` és a `NULL` között.
4. `SQL`-ben is található `NULL`-ok. Mennyiben különböznek az `R`-ben található `NULL`-októl?
5. Hogyan tudjuk ellenőrizni hogy egy objektum `NULL` objektum-e? Ez mikor hasznos?
6. Hozz létre egy `data.frame`-et 3 oszloppal, majd töröld valamelyiket!

Egyéb speciális értékek

1. Milyen egyéb speciális értékeket ismersz `R`-ben?

Kommentek:

1. Milyen céllal hozunk létre kommenteket?
2. Milyen szabványokkal rendelkeznek a kommentek?

Vektorok:

1. Hozz létre üres vektorokat:
 - 4 elemű `character` vektort
 - 10 elemű `numeric` vektort
 - 5 elemből álló `integer` vektort

2. Hozz létre vektorokat a következő értékekkel: (a vektorok nevei legyenek `vec_1`, `vec_2`, ...)
 - 1, 6, 4, 2, 6, 8
 - 1, 2, 3, 4, 5
 - 1-től 1000-ig
 - 3 tetszőleges szót/mondatot
 - 0-tól 1000-ig 50-esével
 - 10 db "Hello" és 20 db "Hi"
 - a, b, c betűket ismételve 100-szor: a, b, c, a, b, c, ..., a, b, c
3. Mi történik, ha:
 - a vektor 0-adik elemét választod ki?
 - -1-edik elemét
 - olyan elemét (indexét), ami nem létezik?
4. Mi történik, ha egy logikai vektorral adom meg a kiválasztandó elemeket?
 - Hogyan befolyásolja az eredményt, ha rövidebb, vagy hosszabb logikai vektort adok meg, mint amiből szeretnék kiválasztani?
5. Válaszd ki az előzőleg készített vektorokból:
 - Az első 3 elemet
 - Az első és utolsó elemeket
 - Utolsó 3 elemet
 - 50-edik elemet
 - A numeric vektorok közül az összes 5-nél nagyobb értékeket
6. Hogyan tudjuk egy vektor bizonyos elemeit módosítani?
7. Tetszőleges numerikus vektorból módosítsd:
 - Az első 5 elemet cseréld le 0-ra
 - Az utolsó elem legyen NA
 - Minden ötödik elemet szorozd meg 2-vel
 - Szúrj be egy elemet a vektor elejére, közepére és végére.
 - Töröld ki az első 3 elemet
8. Milyen információt ad az `str` függvény vektorokra?

Vektorizált műveletek:

1. Számítsd ki a következő műveleteket:
 - `v1 * v2`
 - `v1 + v3`
 - `v2 + v3`
 - `v3 + v4`
2. Mi történik hosszabb/rövidebb vektorok összeadásakor?
3. Miért érdemes R-ben vektorizált műveletekre hagyatkozni?
4. Számítsd ki egy tetszőleges vektor minimumát, maximumát, átlagát, mediánját és szórását. Mi történik ha NA értékeket tartalmaz a vektor?

Listák

1. Mikor érdemes listát használni vektor helyett?
2. Hozz létre egy listát, aminek
 - Az első eleme character
 - 2.ik és 3.ik eleme 1-1 numerikus vektor
3. Lehet névvel és név nélkül is létrehozni a listaelemeket. Hogyan tudjuk ezeket elérni?
4. Válassz ki egy lista elemet index-el (az elem sorszámával)
5. Hozz létre egy listát nevek nélkül, majd utólag adj nevet az elemekhez.
6. Vegyünk egy új listát: `lst <- list(elem1 = 1:5, elem2 = 10:20)`. Módosítsd a listát az alábbiak alapján:
 - Az elem1-ben található vektor elemeihez adjunk hozzá 1-et
 - Az elem2-ben található vektor elemeiből csak a 10 alattiakat tartsuk meg.

Mátrixok

1. Mi a különbség a vektor és a mátrix között? Mi a közös bennük?
2. Hozz létre egy mátrixot `int_mat` néven 1000 elemmel. A mátrix legyen 50x20-as. Oszloponként töltsd fel az értékeket.
3. Hogyan tudjuk ellenőrizni a mátrixot dimenziószámait?
4. Válaszd ki az előzőleg készített `int_mat` mátrixból az első 5 oszlopát.
 - Az eredeti értékeket válaszd ki
 - Minden értéknek a dupláját válaszd ki
 - Az értékeket alakítsd vektorra és csak azokat tartsd meg, amelyek nagyobbak mint 20
5. Válaszd ki a mátrix első sorának első elemét.
6. Oszd el a mátrix első sorának összegét a második sorának összegével.
7. Melyek a mátrix legkisebb, legnagyobb értékei? Menni az átlag, a módusz és a medián?

Mátrixműveletek

1. Mi a különbség a skalárszorzás és a mátrixszorzás között?
2. Mire kell ügyelni mátrixok összeszorzásakor?
3. Lehet-e vektort vektorral szorozni?
4. Szorozd össze az 1, 2, 3, 4 vektort és az 1, 1, 2, 2 vektort úgy, hogy egy 4x4-es mátrix legyen az eredmény.
5. Transzponáld a mátrixot. Mi történik, ha 2-szer transzponálunk egy mátrixot?
6. Készíts 2 db 3 elemű vektort majd fűzd őket össze soronként. Mi történik, ha különböző típusú vektorokat fűzünk össze egy mátrixba?

data.frame-ek

1. A beépített `datasets` csomag több R objektumot, köztük `data.frame`-eket is tartalmaz. Pl. `datasets::cars`. A `cars` dataset-et másoljuk le egy saját R objektumba.
 - Készítsünk egy új oszlopot `dist_per_speed` néven, ami a `dist / speed` értékeket tartalmazza. Pl. első sor: $2 / 4 = 0.5$
 - Nevezzük át a `speed` oszlopot `speed_mph`-ra és hozzunk létre egy új oszlopot `speed_kmh` néven, ami km/h-ban tartalmazza a sebességet, nem mérföld / órában.
 - Ugyanezt megcsinálhatjuk a `dist` oszlopra is, ami láb (ft) mértékegységet használ. Ezt alakítsuk méterre.
 - Készítsünk egy új `data.frame`-et, ami csak az általunk kreált oszlopokat tartalmazza
 - Átlagosan hány mekkora a fékezési távolság a `data.frame`-ben?

Függvények

1. Hozz létre egy függvényt **absolute_value** néven amely két számot kap bemenetként (x és y), majd visszaadja a két szám különbségének abszolútértékét.
2. Hozz létre egy függvényt **head_tail** néven amely egy vektort kap bemenetként és egy számot (n), majd a vektor első és utolsó n elemét adja vissza vektorként.

Pl.:

3. Hozz létre egy függvényt **top_n** néven amely egy vektort kap bemenetként és egy számot (n), majd a vektor n db legnagyobb elemét adja vissza vektorként.
4. Hozz létre egy függvényt **top_bottom** néven amely a következő paraméterekkel rendelkezzen:
 - x: egy vektort
 - top: logikai vektor, alapértelmezetten legyen igaz. Amennyiben igaz, akkor az n db legnagyobb érték kerüljön visszaadásra, ha hamis, akkor az n db legkisebb.
 - n: hány elemet adjon vissza a függvény.
5. Írj egy függvényt **closest_to_median** néven, amely egy numerikus vektort kap bemenetként és visszaszadja a mediánhoz legközelebbi értéket/értékeket
6. Készíts függvényt **longest_char** néven ami egy character vektorból a leghosszabbat adja vissza.

Nehezebb feladatok: A következő feladatokhoz szükséges alap nem szerepelt a tananyagban.

7. Készíts függvényt **roman_to_int** néven amely egy tetszőleges római számot átalakít arab számmá
8. Készíts egy függvényt **word_count** néven, amely egy tetszőleges character vektorban megszámolja a szavak számát.
- b) A **word_count** függvény tartalmazzon egy **unique** nevű logikai vektor argumentumot, ami igaz érték esetén a többször előforduló szavakat csak egyszer számolja.
9. Készíts függvényt **character_count** néven amely egy character vektor összes eleménél megszámolja, hogy hány karakterből állnak

Elágazások

1. Ha egy elágazásban az if ()-ben található feltétel igaz, akkor lefuthat-e az azt követő else if, vagy else ág?
2. Mi a különbség az if és az ifelse között?

Ciklusok

1. Milyen ciklusokat ismersz?
2. Mikor érdemes for ciklust használni?
3. Milyen ciklust érdemes használni az alábbi esetekben:
 - Vektor összes elemére szeretnénk futtatni egy függvényt
 - Lista összes elemére szeretnénk futtatni egy függvényt
 - Mátrix soraira szeretnénk futtatni egy függvényt
 - Mátrix oszlopaira szeretnénk futtatni egy függvényt
 - data.frame összes oszlopára szeretnénk futtatni egy függvényt
 - Két azonos hosszúságú vektor elemeire szeretnénk futtatni egy függvényt

Csomagok

1. Mi a különbség a `library` és a `require` függvények között? Mikor melyiket érdemes használni?
2. Mi a különbség a `::` és a `:::` operátorok között?
3. Mikor nem érdemes betölteni egy csomagot `require`, vagy `library` függvénnyel?