

Řešení následujícího zadání uložte vše do JEDNOHO souboru s příponou ".r" s vaším PRIJMENIM! Soubor následně odevzdejte s předmětem „MDAV_1_SEM_<vase prijmeni>“. Řešení jednotlivých příkladů oddělte od sebe pomocí min 5 řádek a 80 komentářních znaků „#“.

#####

Co je hodnoceno:

- Aby kód byl alespoň „rámcově“ efektivně napsaný.
 - o Tj na místě kde je vhodné použít for cyklus aby byl for cyklus, na místě kde je vhodné použít apply funkci – aby byla apply funkce.
- Vhodně pojmenované proměnné a funkce, „rozumně“ strukturovaný kód.
- U většiny příkladů jsou zmíněné funkce a postupy které máte dodržet, to se samozřejmě taky boduje.
- Zda dospějete k zdárnému řešení

#####

1. Příklad na funkce z rodiny apply a aggregate (5b).

- a. V datasetu mtcars pomocí funkcí z rodiny apply/aggregate spočítejte:
 - Průměrnou hodnotou spotřeby (mpg) pro jednotlivé počty válců (cyl)
 - Průměrnou hodnotu všech ukazatelů v třídění na to zda auto má / či nemá automatickou převodovku (am) pro jednotlivé počty válců zároveň (cyl).

2. Příklad na tzv. „wrapper“ funkci (5b)

- a. Vytvořte funkci která při vložení objektu provede výpočet variačního koeficientu ($sd(x)/mean(x)$).
- b. V případě kdy vloženým objektem bude vektor dat tak vypočte tato funkce variační koeficient pro daný vektor jako celek.
- c. V případě že vstupem bude matice či data.frame tak funkce vypočte variační koeficient pro sloupce daného objektu. Zároveň do objektu který navrací uloží původní pojmenování sloupců (daného data.frame / matice)
- d. V případě že objekt nebude ani matice či vektor tak funkce navrátí objekt NULL a zároveň vypíše do příkazové řádky informaci o nevhodně vloženém objektu.
- e. V rámci vytvoření této funkce by jste měli použít alespoň jednu funkci z rodiny apply a podmínku (IF....) .

3. Vytvořte graf pomocí ggplot s následujícími parametry (5b):

- a. Graf bude vytvořen pro data o kosatcích (dataset iris).
- b. Graf bude na ose x znázorňovat proměnnou Petal.Length a na ose y Petal.Width
- c. Data budou znázorněna formou bodů a velikost bodu bude odpovídat součinu hodnot Sepal.Length a Sepal.Width.
- d. Body budou barevně odlišené dle jednotlivé skupiny kosatců (species), přičemž použité barvy budou následující: #391e52, #b8af23 a #868cc9
- e. Pro každou skupinu bude v grafu zvlášť vykreslena regresní křivka
- f. Graf bude mít odpovídající popisky os x, y, a nadpis a dále i legendy.
- g. Graf rozdělíte dle skupiny kosatců (Species) do samostatných oken, která budou uspořádaná v jednom sloupci

4. Příklad na dplyr (10b)

- a. Použijte data 'gapminder' z library gapminder
- b. Nejprve vytvořte soubor „kontinenty“, který bude obsahovat data pro jednotlivé kontinenty a jejich celkovou velikost populace v pozorovaných letech. Očekávané využití funkcí group_by(), summarise(). Chybějící hodnoty do součtu nezahrnujte.
- c. Do souboru „kontinenty“ dopočtete novou proměnou „svet_pop“, která bude obsahovat celkovou světovou populaci v jednotlivých letech. Očekávané využití funkce mutate()
- d. Nyní vytvořte druhý soubor „Evropa“ který obsahuje jen evropské státy, rok a populaci jednotlivých států. Očekávané využití funkcí filter(), select().
- e. K souboru „Evropa“ připojte informaci o velikosti světové populace ve sledovaných letech ze souboru „kontinenty“. Výsledný soubor bude obsahovat pouze proměnné: country, year, pop, svet_pop. Zároveň bude obsahovat pouze unikátní pozorování. Očekávané využití funkcí left_join(), filter(), select().
- f. Ve sloučeném souboru vypočtete podíl jednotlivých evropských států na světové populaci. Výsledný soubor bude seřazen podle let pozorování (year – od roku 1952 do roku 2007) a podle podílu států na světové populaci (od největšího po nejmenší podíl). Očekávané využití funkcí left_join(), select(), unique(), mutate(), arrange().
- g. Zobrazte prvních 7 pozorování výsledného souboru

5. Lineární optimalizace (10b)

- a. Provozujete hospodu na mýtině. V nabídce máte následující položky za následující ceny:
 - Pivo: 30 Kč (0.5l)
 - Utopenec: 25 Kč
 - Klobása: 40 Kč.
- b. Víte, že o víkendu k vám přiletí návštěva a jelikož máte hospodu zcela prázdnou po stránce zboží tak budete muset nakoupit dostatek zásob. O návštěvě víte, že bude disponovat 2.000,- Kč a je jí jedno co bude jíst a pít. Má ale následující požadavky:
 - Chce mít k dispozici minimálně 10 piv.
 - Chce mít k dispozici minimálně 10 jakýchkoliv jídel.
 - Nesní více než 20 utopenců.
- c. Máte malou motorku, která nad rámec své váhy uvezete jen limitované množství zboží. Víte že daná potravina (+ obalový materiál) váží:
 - Pivo (0.5 l) 600gr
 - Utopenec 150gr
 - Klobása 200gr
 - Celkem uvezte jen 15 kg potravin!!
 - Zároveň víte, že v řeznictví, kde nakupujete klobásy budou mít maximálně 30 klobás na skladě!
- d. Zboží nakupujete za následující ceny:
 - Pivo: 15 Kč
 - Utopenec: 10 Kč
 - Klobása: 12 Kč
- e. Definujte úlohu pro lineární programování a najděte řešení pro maximalizaci zisku. Vše srozumitelně okomentujte a interpretujte výsledek.

6. Webscraping + zpracování dat (15)

Vytvořte skript který ze stránky www.obcepro.cz/obce stáhne všechny údaje o všech obcích za všechny dostupné roky. Data stáhnete do R a následně uložíte podle jednotlivých let do souborů s názvy „OBCE_<rok>.csv“ kde za výraz <rok> dosadíte číslo roku za který dané údaje jsou stažené. V případě kdy v nějakém roce nebudou pro každou obci dostupné všechny údaje, které jsou dostupné za ostatní obce, v daném roce tak přidejte do záznamu informaci o chybějícím pozorování formou NA.

Data budou uložena ve formátu „širokých dat“ kde krom všech dostupných proměnných za daný rok pro danou obci bude navíc odpovídající proměnné: „obce“, „kraj“, „orp“ a „orkes“.

Následně vykreslete tři mapy české republiky pro vybrané tři ukazatele po jednotlivých obcích, kde jednotlivé obce vykreslíte barevnou škálou dle intenzity daných ukazatelů. Pro vykreslení doporučuji využít knihovnu RCzechia anebo obdobný nástroj v jazyku R.