

Domácí úkol č. 3

Zadání

V této úloze máte upravit R skript `hw03.R`. Svůj kód zadejte vždy těsně pod zadání každého úkolu. Neupravujte žádné jiné řádky, zejména ne řádky, které volají funkce `load()` a `save()`. Samozřejmě však můžete na vhodná místa vložit další řádky kódu.

Vášim úkolem je spočítat, jak se meziročně měnil v procentech objem dopravy ve vybrané zemi. Bohužel však máte pouze data o podílu a objemu vlakové dopravy. Datový soubor “`hw03.RData`” obsahuje dva datasety třídy *tibble* (zdroj dat je Eurostat):

- `modal_split` uvádí údaje o podílech jednotlivých módů dopravy od roku 1993 do 2015; má následující sloupce:
 - `unit` ... jednotky (vždy PC, tj. procenta)
 - `vehicle` ... mód dopravy (CAR znamená osobní auta, TRN vlaky a BUS_TOT různé autobusy)
 - `geo` ... kód země (zde vždy CZ, tj. Česká republika)
 - `time` ... první den v daném roce (důležitý je reálně rok)
 - `values` ... podíl daného dopravního módu v procentech; 10.2 znamená, že daný mód má v daném roce 10.2 % na celkovém objemu dopravy v zemi
- `train_total` uvádí údaje o celkovém objemu železniční dopravy od roku 2004 do 2015; má následující sloupce:
 - `unit` ... jednotky (MIO_PKM jsou miliony pasažéro-kilometrů, THS_PAS nás nezajímá)
 - `tra_cov` ... vždy TOTAL
 - `geo` ... kód země (jako výše)
 - `time` ... čas (jako výše)
 - `values` ... hodnota proměnné; v řádcích, kde je `unit` roven MIO_PKM, je to objem milionů pasažéro-kilometrů ujetých v dané zemi v daném roce vlakem

Přitom je zaručeno, že data obsahují všechny údaje pro všechny roky z daného intervalu a že data jsou uspořádána podle času od nejnovější hodnoty po nejstarší.

Váš cíl je postupně:

1. eliminovat z datasetu `modal_split` všechny dopravní módy mimo vlaku (tj. necháte pouze `vehicle` TRN) a všechna pozorování starší než 1. 1. 2004 (tj. necháte pouze pozorování `time` `>= "2004-01-01"` – použijte toto jako kus kódu) (2 body)

2. eliminovat z datasetu `train_total` všechna pozorování kromě těch, kde je jednotka MIO_PKM (2 body)
3. vytvořit nový dataset třídy *tibble*, který pojmenujete `transit1`; bude obsahovat postupně tyto sloupce:
 - `time` ... čas (jako v datech)
 - `train_share` ... podíl vlakové dopravy z datasetu `modal_split`
 - `train_volume` ... objem vlakové dopravy z tabulky `train_total`
 - `transit_volume` ... objem *veškeré* dopravy v daném roce v zemi (spočítáte z objemu železniční dopravy a podílu vlaků na veškeré dopravě) (2 body)
4. do datasetu `transit2` uložit kopii datasetu `transit1` uspořádnou podle času vzestupně, tj. tak že první řádek bude pro rok 2004, druhý 2005 atd.; pozor! dataset `transit1` se při tom nesmí změnit! (2 body)
5. do datasetu `transit3` uložit kopii datasetu `transit2`; pak do `transit3` přidat proměnnou `transit_growth`, která bude obsahovat meziroční tempo růstu veškeré dopravy v zemi; spočítáte je jako $(y_t/y_{t-1}-1) \times 100$, tj. 10.2 znamená 10.2 %; pozor vektor temp růstu je kratší než ostatní vektory, takže musíte na vhodná místa umístit NA; pozor! datasety `transit1` a `transit2` se při tom nesmí změnit! (2 body)

Doporučení

Přečtěte si důkladně dokument *Jak zpracovat domácí úkol* v ISu.

Kontrola

V tomto domácím úkolu mimořádně uvádíme i část správných výsledků na testovacím vzorku dat. Růst celkové dopravy v jednotlivých letech uvádí tabulka:

time	train_share	train_volume	transit_volume	transit_growth
2004-01-01	7.5	6580	87733.33	NA
2005-01-01	7.3	6667	91328.77	4.0981388
2006-01-01	7.5	6922	92293.33	1.0561472
2007-01-01	7.3	6898	94493.15	2.3835062
2008-01-01	7.1	6773	95394.37	0.9537363
2009-01-01	6.8	6472	95176.47	-0.2284156
2010-01-01	7.5	6559	87453.33	-8.1145447
2011-01-01	7.6	6669	87750.00	0.3392285
2012-01-01	8.3	7196	86698.80	-1.1979542
2013-01-01	8.5	7512	88376.47	1.9350620
2014-01-01	8.4	7644	91000.00	2.9685836

time	train_share	train_volume	transit_volume	transit_growth
2015-01-01	8.6	8125	94476.74	3.8205980

Poznámky

1. V této úloze máte postupně vytvořit proměnné `transit1`, `transit2` a `transit3`. To je jen proto, abyste pozdější manipulací nezneškodili předchozí výsledky. Normálně byste jen postupně upravovali jedinou proměnnou `transit`.
2. Některé z operací, které jste v této úloze prováděli, vás později naučíme provádět mnohem elegantněji s pomocí funkcí z balíku **dplyr**.