

Domáca úloha 2

V súbore `poistky.txt` sú údaje o poistných udalostiach:

- `charge` vyjadruje veľkosť škody pre danú udalosť,
- `region` vyjadruje región USA, kde k poistnej udalosti došlo.

Niektorým udalostiam však chýba údaj o výške škody. Úlohou je pre každý región zistiť, koľko chýbajúcich meraní máme.

1. Vypočítajte počet chýbajúcich meraní po regiónoch pomocou `for` cyklu, ktorý ide od 1 po počet udalostí – a v každom kroku sa pozrie, či je meranie chýbajúce. Ak je chýbajúce tak sa pozrie, do ktorého regiónu udalosť patrí, a podľa toho ho započíta.
2. Počet chýbajúcich meraní po regiónoch vypočítajte aj pomocou `for` cyklu, ktorý ide iba po regiónoch (teda cca. od 1 po 4) – v každom kroku nájde všetky udalosti prislúchajúce tomu regiónu a určí počet chýbajúcich meraní spomedzi nich (efektívnym spôsobom, bez ďalšieho `for` cyklu).
3. Overte, či postupy 1 a 2 dali ten istý výsledok. Zmerajte tiež čas trvania postupov 1 a 2 (nie iba raz, ale vypočítajte napr. priemernú dobu trvania na základe 100 opakovaní). Sformulujte ponaučenie pre čitateľa: ktorý postup je rýchlejší a prečo?
4. *Bonus (2 body)*: Výpočet spravte aj najviac “elegantne” pomocou funkcie `tapply` alebo podobnej funkcie. Porovnajte výstupy a rýchlosť s predchádzajúcimi postupmi.

Pomôcky:

- Dáta sa dajú ľahko načítať pomocou `read.table`. Nezabudnite nastaviť, že údaje sú oddelené pomocou “;” a že súbor obsahuje header (viď `help` k `read.table`: voľba `sep` a `header`).
- Zoznam regiónov v datasete vieme získať napríklad pomocou `unique` (pozrite `help` k tej funkcii).
- Funkcia `is.na` pre vstupný vektor x určí, ktoré jeho prvky sú NA (jej výstupom je vektor hodnôt TRUE/FALSE, kde TRUE je na tých pozíciach, kde x má hodnoty NA).