Zápočtový úkol

Jan Caha ([jan.caha@mendelu.cz](mailto:jan.caha@mendelu.cz))

24. 4. 2019

Cílem zápočtového úkolu je zpracování vámi zvolené datové sady v R, za pomoci postupů a metod prezentovaných na cvičení. Výsledkem bude projekt pro RStudio (adresář), který bude obsahovat data (ve vhodné složce) a jeden R Markdown soubor, v rámci něhož provedete načtení, zpracování a prezentaci data dle detailů uvedených dále.

Vyberte si libovolnou datovou sadu, např. některou dostupnou z [Národního katalogu otevřených dat](https://data.gov.cz/datov%C3%A9-sady), ale lze zvolit i jiná data. Výběr datové sady je zcela na vás, datový sada ani nutně nemusí být prostorová data, byť by to bylo vhodnější. Předpokládám, že každý si vybere jinou datovou sadu, či případně, pokud by došlo ke shodně ve volbě datových sad, tak nebude shodný postup zpracování. V opačném případě by nebylo možné udělit zápočet za předmět.

Výsledný RMarkdown se bude překládat do html dokumentu a bude obsahovat alespoň 2 grafy, různých typů - používajících různé geom\_\*, vytvořené pomocí ggplot2 a jednu vizualizaci prostorových dat vytvořenou pomocí balíku tmap. Pakliže zvolíte data neprostorová, pak je nutné vytvořit 4 grafy. Dále bude v dokumentu alespoň jedna interaktivní tabulka (se stránkováním, aby nebyla příliš dlouhá). Nedílnou součástí dokumentu bude i veškerý zdrojový kód použitý na zpracování dat.

Předpokládá se, že práce s daty bude obsahovat alespoň několik dílčích kroků, filtrování, joinovaní atd. Jako příklad lze použít ukázku z prezentace *Geografická data v R* týkající se dat z voleb. Není možné striktně stanovat pravidla pro zpracování dat, ale předpokládá se, že postup, který zvolíte nebude pouze triviální načtení a vizualizace, ale že bude obsahovat i další dílčí kroky.

Výsledný projekt zabalte do souboru .zip a odešlete na mail uvedený výše. V případě, že by byl problém s velikostí souboru, použijte libovolné úložiště a pošlete odkaz.