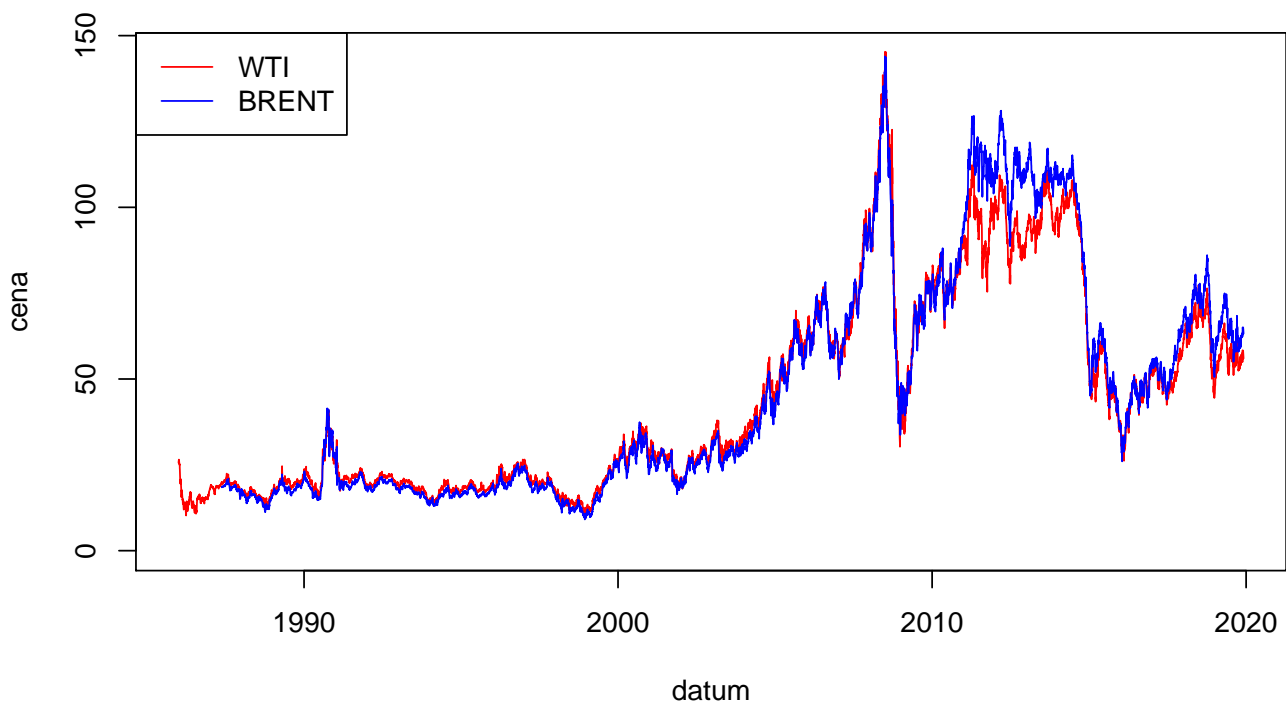


M4130 cvičení 13

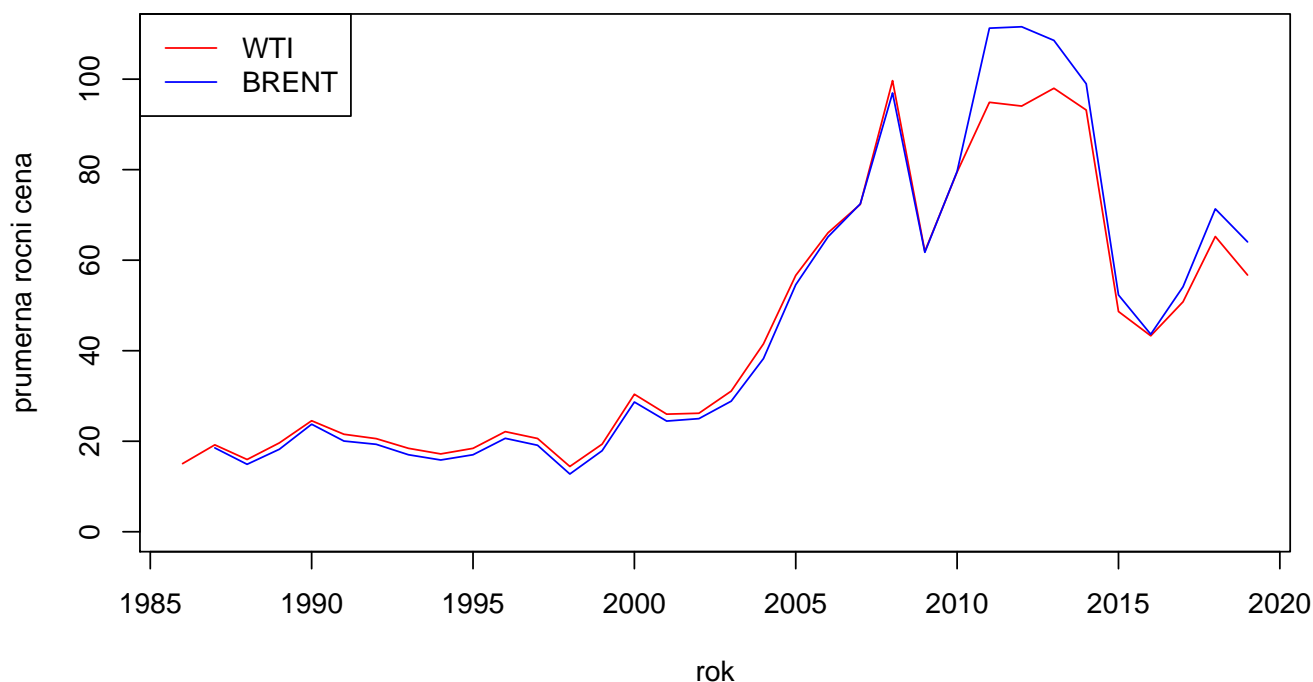
Analýza dat v R

9. 12. 2019

1. **Ceny ropy.** Ze souboru `oil.csv` načtěte do datové tabulky `X` cen ropy dvou typů (WTI a BRENT) za období 2. 1. 1986 – 2. 12. 2019 pomocí funkce `read.csv`: soubor má hlavičkový řádek, používá oddělovač buněk ; a desetinnou čárku. Pomocí funkce `ISOdate` do tabulky doplňte sloupec s datem v ISO formátu `YYYY-MM-DD`. Do jednoho grafu vykreslete průběh cen obou druhů ropy v závislosti na datumu.

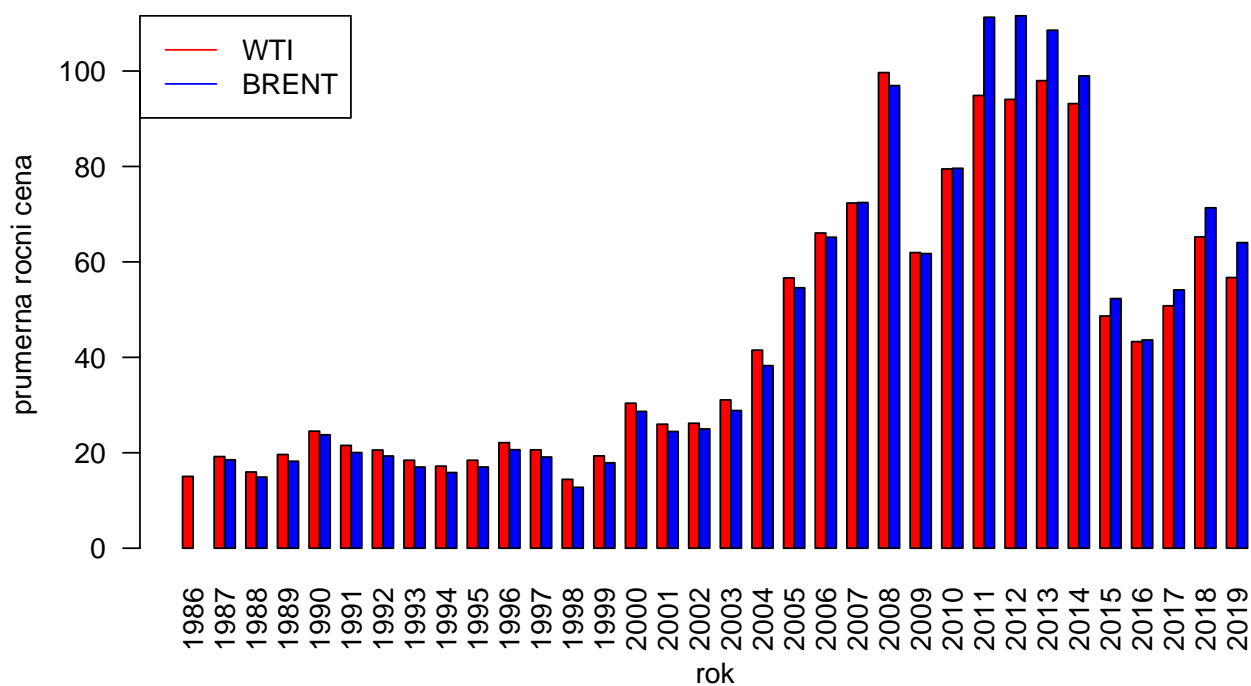


Pro každý typ ropy vytvořte tabulku (pomocí `tapply`), v níž budou spočítané průměrné roční ceny ropy pro každý rok. Průměrné roční ceny pro oba typy ropy následně vykreslete do jednoho grafu.



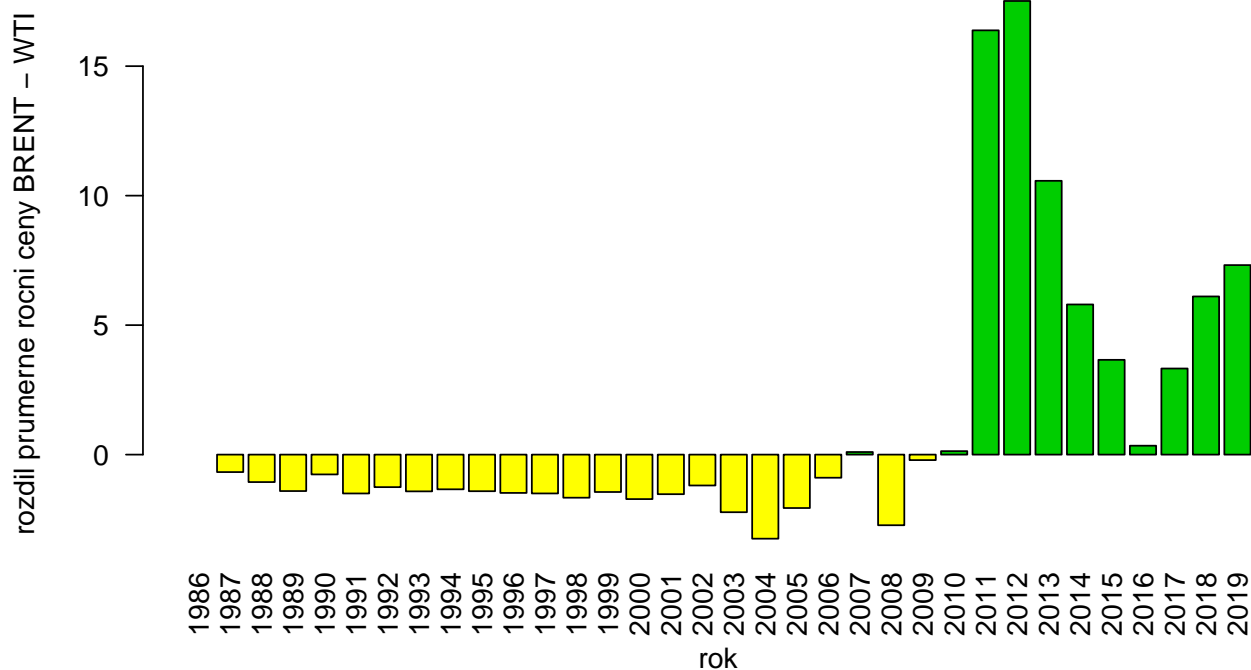
Obě tabulky ročních průměrných cen slučte (`merge`) do jedné tabulky `Y` se sloupci `cena`, `rok` a `typ`. Tabulku `Y` využijte k vytvoření matice `Z`, z níž vykreslete následující sloupcový graf.

```
Z <- tapply(Y$cena, list(Y$rok, Y$typ), mean)
```



Pomocí `Z` a funkce `apply` s vhodně definovanou vlastní funkcí spočítejte rozdíly ročních průměrných cen ropy `BRENT-WTI` a vykreslete je do sloupcového grafu. Kladné a záporný rozdíly přitom

vykreslete různými barvami.



2. **Titanic.** Ze souboru `titanic1.csv` načtete do datové tabulky `X` údaje o pasažérch *Titanicu*.

Popis dat: `pclass` = cestovní třída (1, 2, 3), `survived` = přežili? (0 = ne, 1 = ano), `sex` = pohlaví (female, male), `age` = věk, `fare` = cena lístku, `embarked` = nástupní město (C = Cherbourg, Q = Queenstown, S = Southampton).

Nečíselné proměnné (vše kromě `age` a `fare`) příp. opravte na faktory. Zároveň upravte úrovně `pclass` na `class 1 / class 2 / class 3`, `survived` na `no / yes` a `embarked` na `Cherbourg / Queenstown / Southampton`.

Spočítejte tabulku počtu pasažérů v rozdělení podle cestovní třídy a pohlaví. Spočítejte z ní, kolik bylo celkem žen/můžů a cestujících v 1./2./3. třídě.

Spočítejte tabulku počtu přeživších pasažérů v rozdělení podle cestovní třídy a pohlaví. Spočítejte tabulku procent (zaokrouhlení na 1 desetinné místo) přeživších v rozdělení podle cestovní třídy a pohlaví. Počty pasažérů a počty přeživších vykreslete i pomocí sloupcových grafů (pohlaví barevně odlište).

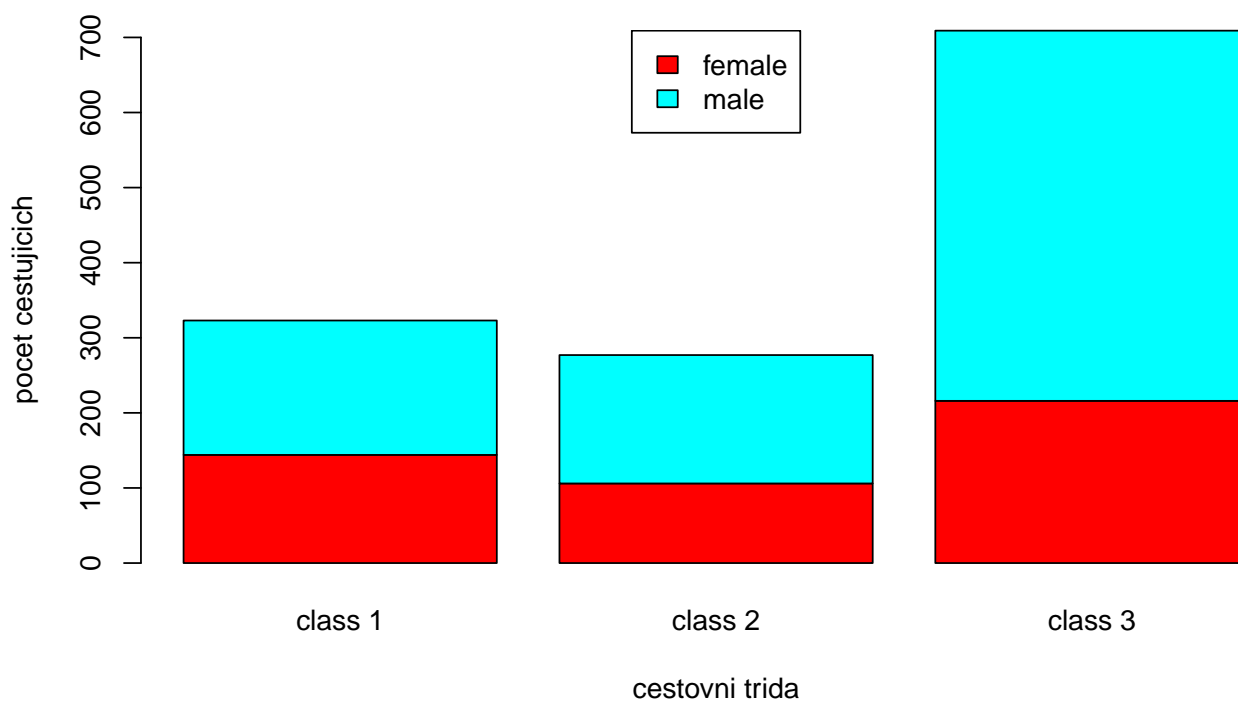
	female	male
class 1	144	179
class 2	106	171
class 3	216	493

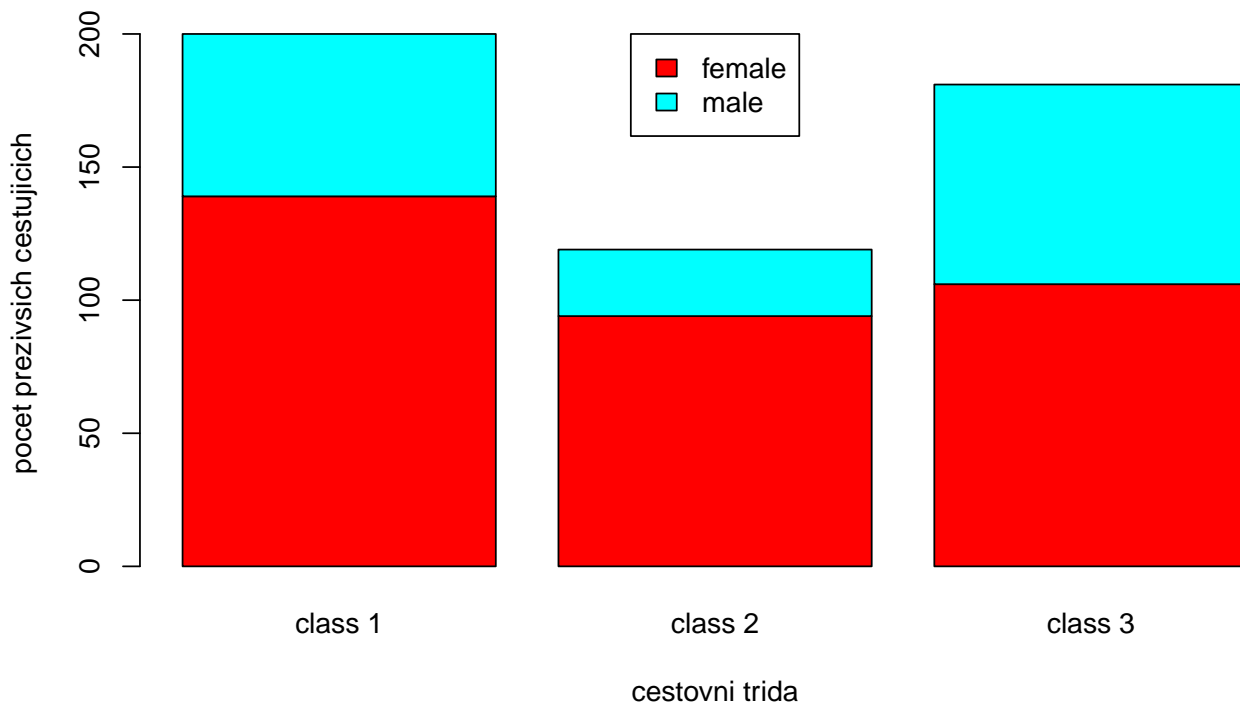
female	male
466	843

class 1	class 2	class 3
323	277	709

	female	male
class 1	139	61
class 2	94	25
class 3	106	75

	female	male
class 1	96.5	34.1
class 2	88.7	14.6
class 3	49.1	15.2





Spočítejte tabulku průměrných věků pasažérů (zaokrouhlení na 1 desetinné místo) v rozdělení podle cestovní třídy a pohlaví. Analogickou tabulku průměrných věků spočítejte jen pro přeživší.

```

      female male
class 1  37.0 41.0
class 2  27.5 30.8
class 3  22.2 26.0

```

```

      female male
class 1  37.1 36.2
class 2  26.7 17.4
class 3  20.8 22.4

```

Spočítejte tabulku průměrných cen a tabulku maximálních cen lístků pasažérů (zaokrouhlení na 1 desetinné místo) v rozdělení podle cestovní třídy a pohlaví.

```

      female male
class 1 109.4 69.9
class 2  23.2 19.9
class 3  15.3 12.4

```

```

      female  male
class 1  512.3 512.3
class 2   65.0  73.5
class 3   69.5  69.5

```

Vytvořte tzv. *čtyřpolní tabulku* závislosti přežití na pohlaví: tabulka uvádí počty cestujících v rozdělení podle přežití (řádky) a pohlaví (sloupce).

	female	male
no	127	682
yes	339	161

Spočítejte tzv. *šanci* na přežití, tj. počet přeživším podělený počtem nepřeživších, a to zvlášť pro ženy a pro muže. Nakonec spočítejte tzv. *poměr šancí* pro ženy a pro muže, což je šance pro ženy podělená šancí pro muže.

[1] 2.669291

[1] 0.2360704

[1] 11.30718