

4. Domaća zadaća - ROVKP

Vinko Kolobara
4. lipnja 2017.

1 ZADATAK: RAD S KOLEKCIJSKIM TOKOVIMA

Koliko je bilo ulaznih datoteka senesorscope-monitor-xx.txt?
97

- Koliko se zapisa nalazi u izlaznoj datoteci?
4726643

- Kolika je veličina izlazne datoteke?
407.800.597 byte-a

2 ZADATAK: OBRADA PODATAKA PROGRAMSKIM OKVIROM APACHE SPARK

1. Koje je najnepopularnije žensko ime kroz čitav period i države?

Anaissa

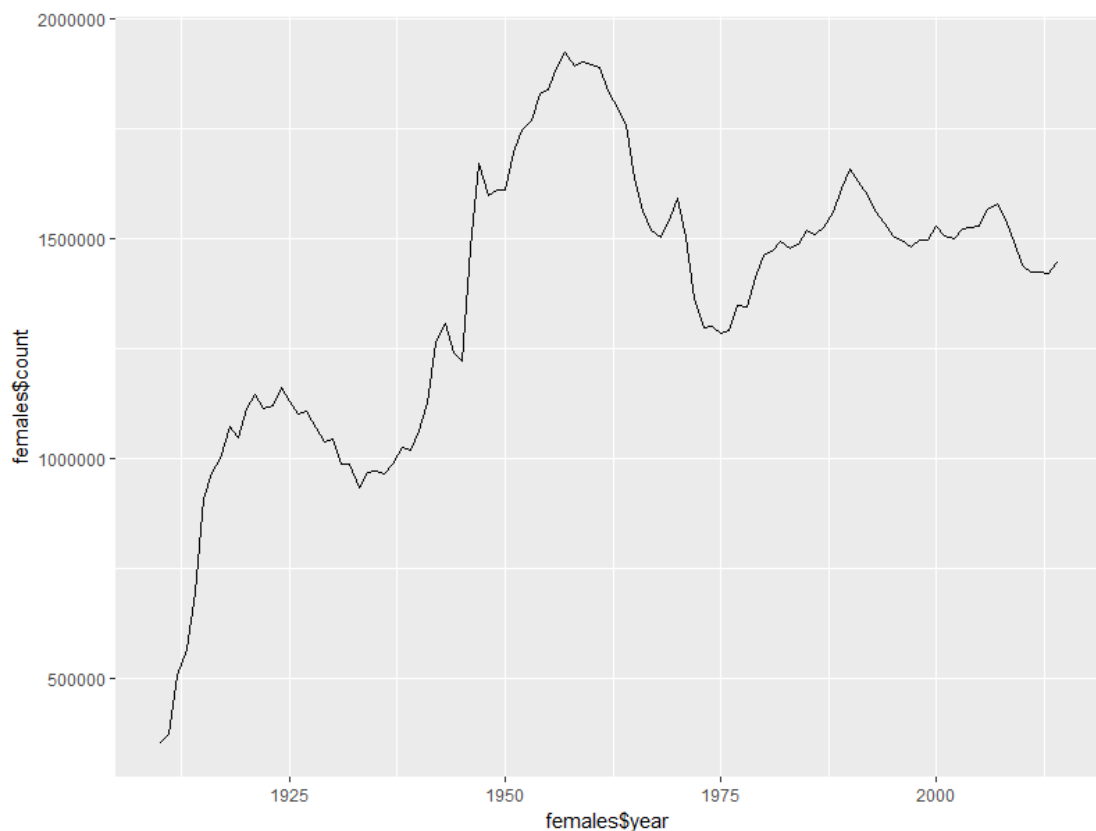
2. Kojih 10 muških imena su najpopularnija kroz čitav period i države?

James, John, Robert, Michael, William, David, Richard, Joseph, Charles, Thomas

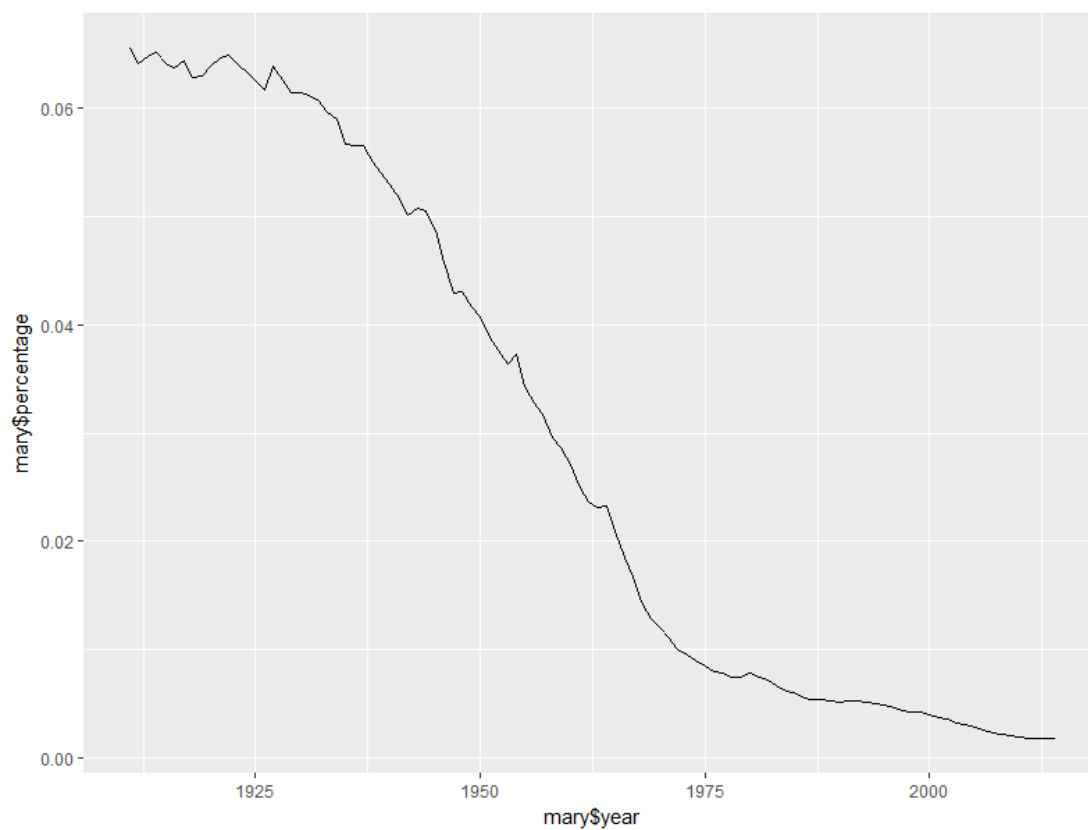
3. U kojoj državi je 1946. godine rođeno najviše djece oba spola?

NY

4. Kakvo je kretanje broja novorođene ženske djece kroz godine? Rezultat je (sortirana) lista tipa Pair2 (ključ je godina, a vrijednost je broj novorođenčadi)



5. Kakvo je kretanje postotka imena Mary kroz godine? Rezultat je (sortiran) skup tipa Pair2 (ključ je godina, a vrijednost je postotak). Pri tome iskoristite polje iz prethodnog pitanja.



6. Koji je ukupni broj rođene djece u cjelokupnom periodu u svim državama?
298883326

7. Koliki je broj različitih imena koja se pojavljuju u zapisima?
30274

3 ZADATAK: OBRADA TOKA PODATAKA PROGRAMSKIM OKVIROM APACHE SPARK

Koliko često (u sekundama) nastaje novi direktorij na disku?
Svakih 10 sekundi.

Koliko često (u sekundama) se pokreće izračun?
Svakih 10 sekundi.

Može li vrijednost parametra `solarPanelCurrent` neke stanice biti manja u nekom direktoriju nego u njegovom neposrednom prethodniku. Zašto?

Može, svaki sljedeći direktorij za izračun uzima zadnjih 60 sekundi (svakih 10 sekundi), tako da je moguće da se maksimum prethodnog nalazio u 10 sekundi koje su prethodile trenutnom prozoru.

Kako se kreću vrijednosti parametra `solarPanelCurrent` neke postaje u prva 3 direktorija koji su nastali? Zašto?

Brzo se povećava dok ne dosegne neku uobičajenu vrijednost, a razlog tome je vjerojatno što se tada tek uključio senzor i uređaj na kojem mjeri.