

Hemuppgift 3

R för samhällsvetare. Stockholm universitet. Sommar 2020.

Besvaras senast måndag 6 juli 17:00.

1. Inlämningsuppgifter besvaras genom att ladda upp ett R-skript på Athena . Om du föredrar går det även bra att redovisa dina svar i ett worddokument, pdf, powerpoint eller dylikt. Döp filen till "*Hemuppgift_3.R*". Du behöver inte ange ditt namn.
2. Det väsentliga är inte att dina resultat är snyggt presenterade. Huvudsaken är att du finner någon form av svar på frågan. Oavsett filformat ska det tydligt framgå vilka kommandon du använder för att lösa respektive uppgift. Skriv kommentarer i din kod (#).
3. Om du kör fast på någon fråga – lämna inte blankt. Gör så gott du kan och förklara kort vad som inte fungerar.

Uppgift

1. Importera följande fyra böcker i R:
 - Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*
 - Malthus, *An Essay on the Principles of Population*
 - Ricardo, *On the Principles of Political Economy, and Taxation*
 - Marx, *A Contribution to The Critique of the Political Economy*

Gör det antingen med hjälp av .rds-filen som finns bifogad på Athena eller med följande kod:

```
install_packages("gutenbergr")
library("gutenbergr")
valfritt_namn <- Gutenberg_download(c(3300, 46423, 33310, 4239), meta_field = "title")
```

2. Tokenisera texten. Varje analysenhet ska vara ett ord och endast små bokstäver.
3. Importera lista med stoppord. En lista med engelska stoppord finns uppladdad på Athena. Samma lista kan nås direkt i R med koden
`stopwords <- tidytext::stop_words`
4. Rensa all data från stoppord.
5. Ta fram en lista på de 20 vanligaste orden *i respektive bok*.
6. Illustrera resultatet från fråga 5 med hjälp av ett eller flera diagram.