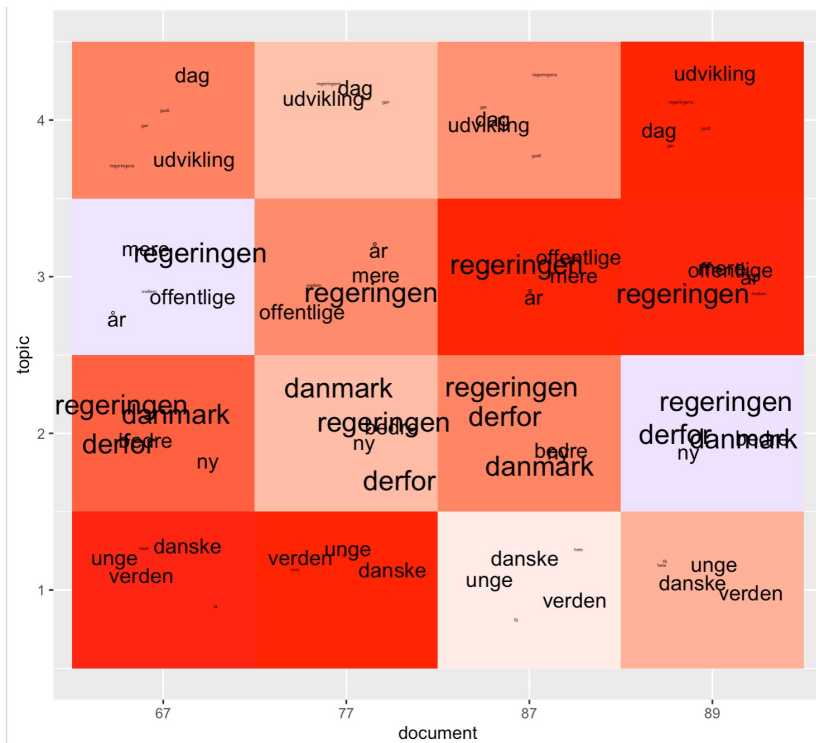


Opgave 98



Indlæs rds-filen med danske åbningstaler.

Dan variabelen "year" fra url og/eller titel.

Dan variabelen "speaker" vha filen "sp.rds".

Ikke alle taler er fra folketingets åbning. Lav derfor en variabel "Folketing" som kan bruges som filter.

Dan variabelen "score" som skal være en sentida-sentimentscore.

Vælg fire taler som dækker bredt politisk. Tokenize til ord og lav en count-dataframe på tf samt tf-idf. Du skal samtidig lave en sentimentscore beregnet ud fra filen "dsl.csv".

Lav dine data om til en document term matrix og kør en LDA med 4 topics.

Vurder om hver tale har haft sit eget tema og plot

Illustrer alle fire taler med en wordcloud på top 20 tf

