Code for knitting

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

Code for installing and loading relevant packages

install.packages("tidyverse")

install.packages("tidytext")

install.packages("lubridate")

library(tidyverse)

library(tidytext)

library(lubridate)

Code for loading data - CSV.file

andersen\_data <- read\_csv("http://labs.statsbiblioteket.dk/labsapi/api/aviser/export/fields?query=%22h%20c%20andersen%22%20AND%20py%3A%5B1855%20TO%201875%5D%20AND%20pu%3A%28%C3%A5rhus%20OR%20odense%29&fields=link&fields=recordID&fields=timestamp&fields=pwa&fields=cer&fields=fulltext\_org&fields=pageUUID&fields=editionUUID&fields=titleUUID&fields=editionId&fields=familyId&fields=newspaper\_page&fields=newspaper\_edition&fields=lplace&fields=location\_name&fields=location\_coordinates&max=-1&structure=header&structure=content&format=CSV")

Code for extracting the years from the timestamp column:

andersen\_data %>%

mutate(y = year(timestamp)) -> andersen\_data

Code for counting occurrances and visualization:

andersen\_data %>%

count(y, sort = FALSE) -> andersen\_count

Code for visualization with geom\_line:

andersen\_count %>%

select(n, y) %>%

ggplot(aes(x = y, y = n)) +

geom\_line() +

labs(title = "Newspapers containing H.C. Andersens name") +

xlab("Year") +

ylab("Newspaper count") +

theme(axis.text.x = element\_text(angle=90))

Code for visualizing with reorder function:

andersen\_count %>%

select(n, y) %>%

ggplot(aes(x = reorder(y,n), y = n)) +

geom\_col() +

labs(title = "H.C. Andersen name occurance in newspapers sorted by occurance") +

xlab("Year") +

ylab("Name count") +

theme(axis.text.x = element\_text(angle=90))

Code for tidying the data:

andersen\_data %>%

unnest\_tokens(word, fulltext\_org) -> andersen\_tidy

andersen\_tidy %>%

count(y,word, sort =TRUE)

Code for stopwordlist and implement:

stopword\_1800 <- read\_csv("https://gist.githubusercontent.com/maxodsbjerg/1537cf14c3d46b3d30caa5d99f8758e9/raw/9f044a38505334f035be111c9a3f654a24418f6d/stopord\_18\_clean.csv")

andersen\_tidy %>%

anti\_join(stopword\_1800) %>%

count(y,word, sort =TRUE)

andersen\_tidy %>%

count(y,word, sort =TRUE) %>%

filter(word %in% c("eventyr"))

Code for visualizing the connection between words:

andersen\_tidy %>%

count(y,word, sort =TRUE) %>%

filter(word %in% c("eventyr", "andersen")) %>%

ggplot(aes(x = y, y = n, color = word)) +

geom\_line() +

labs(title = "H.C. Andersen related words occurance in Odense and Århus newspapers") +

xlab("Year") +

ylab("Word count") +

theme(axis.text.x = element\_text(angle=90))