Citizenship in the years 1740-1862

Rikke Grønning Sell Pedersen

3/12/2019

## Index

* Introduction
  + Packages
  + Collection of data
  + Reading of data
* Visualization

## Introduction

The purpose of this assignment has been to analyze a dataset through the use of the digital methods aquired through the course of the same name by Adela Sobotkova. Each step in the analysis has been carefully noted, to ensure the best possible circumstances for reproducibility.  
The tool used in this assignment is RStudio in version 3.6. It is vital for any attempt to reproduce the outcome of the assignment, that the same version of RStudio is used.

The entire project has been run through an R-Markdown document, which was knitted to a word document to enable the wordcount function.

### Packages

Prior to starting my analysis of a given dataset, I have installed the necessary packages. In this case the necessary packages are all included in the tidyverse library. The [tidyverse](www.tidyverse.org) is “an opinionated collection of R packages designed for data science” (henvisning til tidyverse), and is a helpful way of visualizing data in RStudio. The library is a collaboration of code made by different programmers. Initially, it was only possible to download these packages one by one, but as the community of programmers discovered that there was an outspoken need for a combination-package with all the different packages included, they went together to create the tidyverse libray. This means that we are working in an open source library, and there are several different “authors”. Working with open sources such as these would not usually be considered trustworthy according to most typical methods used by historians, but as working with programming is an entirely different method, the same considerations do not necessarily apply to this case. Having many different authors on a library like this means that it has been run by many different people several times, which heigthens the chances of small errors to be discovered. This might actually mean that an opensource in digital methods can sometimes be a form of quality control.

library(tidyverse)

## -- Attaching packages --------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

### Collection of data

After the installation of the tidyverse, I searched for a dataset online. Through a search in Github, I found a dataset from Aarhus Stadsarkiv, regarding records of citizenships granted in the years 1740-1862. citizenship/1740-1862/citizenship-records-1740-1862-original.csv

I downloaded the csv files directly from github, by pressing the file called citizenship-records-1740-1862-original.csv. I then pressed the “raw” button in the top right corner, and copied the URL. Because I had already run the tidyverse library, I typed the command with an underscore (\_) instead of a dot (.), as the underscore often replaces the dot when working in tidyverse. I named my dataset Citizenships, and the command then ended up looking like this: Citizenships <- read\_csv(“<https://raw.githubusercontent.com/aarhusstadsarkiv/datasets/master/citizenship/1740-1862/citizenship-records-1740-1862-original.csv>”) (Something about Aarhus Stadsarkiv)

Citizenships <- read\_csv("https://raw.githubusercontent.com/aarhusstadsarkiv/datasets/master/citizenship/1740-1862/citizenship-records-1740-1862-original.csv")

## Parsed with column specification:  
## cols(  
## arkiv = col\_character(),  
## bye = col\_character(),  
## herred = col\_character(),  
## amt = col\_character(),  
## sognenr = col\_double(),  
## recnr = col\_double(),  
## folie = col\_character(),  
## aar = col\_double(),  
## borgerskabsdagen = col\_character(),  
## borgered = col\_character(),  
## borgerskabsdato = col\_character(),  
## fornavn = col\_character(),  
## efternavn = col\_character(),  
## oprindelsessted = col\_character(),  
## land = col\_character(),  
## alder = col\_character(),  
## hovederhverv = col\_character(),  
## noter = col\_character()  
## )

### Reading of data

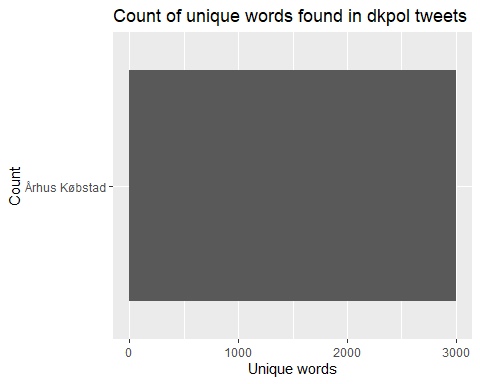
The analysis of the dataset will take its stand in the visualisation of the dataset. As I have no prior knowledge of the information it contains, the visualisation will work as a reading of the data, and the analysis will hereafter consist of contextualizing the data into the historical context. My sources for the historical context are the readme file from Aarhus Stadsarkiv’s Github account where I found the dataset to begin with. The readme file does not, however, give a broader understanding of the time period, so for this, and for a closer understanding of some of the laws implemented within the timescape of the data, I have used [danmarkshistorien.dk](https://danmarkshistorien.dk/leksikon-og-kilder/vis/materiale/naeringsfrihedsloven-1857/)

## Visualization

With my dataset in place in RStudio, I can now begin the visualization of the data in order to analyse it. For my visualization, i have chosen not to embed on the long journey of creating my own code, and have instead found a usable code created by [Max Odsbjerg Pedersen](http://hax.odsbjerg.dk/twitter_scrape.html). The original example for this code uses a twitterscrape and analysis of the #dkpol. The example set up by Max largely follows a set of basic instructions in textmining from [EarthLab](https://www.earthdatascience.org/courses/earth-analytics/get-data-using-apis/text-mining-twitter-data-intro-r/), and can be modified to use on datasets that are not related to twitter. My initial visualization was to see, if the dataset consists of citizenships granted in other cities than Aarhus. This information could also have been aquired by a close reading of the readme file, but for a quicker way to answer this question, I used the code by Max, and replaced one of the originally searched words with the word “bye” (city).

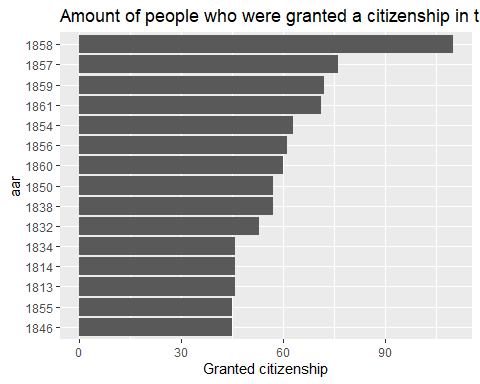
Citizenships %>%  
 count(bye, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(bye = reorder(bye, n)) %>%  
 ggplot(aes(x = bye, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "Count",  
 y = "Unique words",  
 title = "Count of unique words found in dkpol tweets")

## Selecting by n

   
As expected, the only city this dataset contains information about, is Aarhus, so it does not make sense to go any further with this particular thread.  
To gain a deeper understanding of the context of the data, I tried to look at the years in which most citizenships were granted. This would hopefully give me a hint to some historical aspects that might have had an effect on the data.

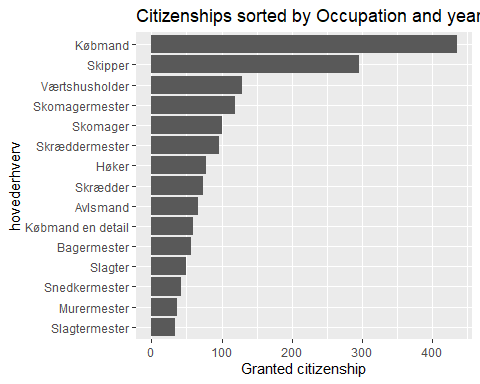
Citizenships %>%  
 count(aar, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(aar = reorder(aar, n)) %>%  
 ggplot(aes(x = aar, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "aar",  
 y = "Granted citizenship",  
 title = "Amount of people who were granted a citizenship in the given year")

## Selecting by n

   
As the graph implies, the year in which most new citizenships were granted in Aarhus, was 1858. There was a significant spike in citizenships compared to the year before, which leads to wonder if anything happened that would have changed the politics on the subject in Aarhus. The law on freedom of trade was decided in December 1957 and implemented from April 1858. The law allowed anyone to posses any occupation, which meant that the monopoly that had been practiced previously - especially by the community of artisans and tradesmen - was dissolved. This might have meant that more people from the countryside wanted to try out an occupation in the city In 1857, the law of (Næringsfrihedsloven)

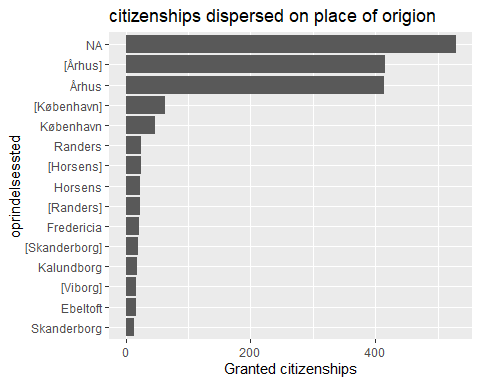
Citizenships %>%  
 count(hovederhverv, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(hovederhverv = reorder(hovederhverv, n)) %>%  
 ggplot(aes(x = hovederhverv, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "hovederhverv",  
 y = "Granted citizenship",  
 title = "Citizenships sorted by Occupation and year")

## Selecting by n

 After 1857, the citizenships were granted primarily to

Citizenships %>%  
 count(oprindelsessted, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(oprindelsessted = reorder(oprindelsessted, n)) %>%  
 ggplot(aes(x = oprindelsessted, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "oprindelsessted",  
 y = "Granted citizenships",  
 title = "citizenships dispersed on place of origion")

## Selecting by n

   
mange hvor det ikke står nogle er anført i [] andre er ikke.. hvorfor? spørg readme.

Citizenships %>%  
 filter(oprindelsessted == "Randers") %>%   
 count(aar, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(aar = reorder(aar, n)) %>%  
 ggplot(aes(x = aar, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "aar",  
 y = "Granted citizenship",  
 title = "Amount of people who were granted a citizenship in the given year")

## Selecting by n

