Citizenship in the years 1740-1862

Rikke Grønning Sell Pedersen

3/12/2019

## Index

* Abstract
* Keywords
  + 1. Introduction
    2. Problems and background
    3. Software framework
    - 3.1 Software Architecture/Prerequisites
    1. Data Acquisition and Processing
    2. Implementation and Empirical Results
    3. Critical Evaluation
    4. Conclusions
* References
* B- Required Metadata: Current executable software version

## Abstract

## 1. Introduction

**Introduce the motivation of developing the script, and explain why it is important (please use language accessible to non-specialized but intelligent audience)**

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 outcome of the analysis has relevance to the future users of Aarhus Stadsarkiv, as it takes an otherwise everlasting dataset, and conforms the informations in it in a more accessible way. This is done through visualisation of the data on the basis of a number of questions put down by the author of this assignment.

## 2. Problems and Background

**Give the formulations of (historical, social, etc.) problems to be solved by the script/software/toolbox** **Introduce the background and related work in literature (cite or list algorithms used, other scripts and software etc.)** This assignment deals with a dataset containing information about the citizenships that were granted in Aarhus from 1740 to 1862. The delimitation of years is based on two other datasets from the same archive, in order to make it easier to work with those three datasets at the same time. The datset contains informations such as name, age, occupation, origin etc. from the applying citizen.

## 3. Software Framework

### 3.1 Software Architecture/Prerequisites

**Give a short overview of the overall software architecture, dependencies and prerequisites**

My computer uses the windows 10 system. 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. As it will occur in the version of the file that I have uploaded to Github, some of the layout has been made directly in R, but due to my limited skills in regular expressions, I did the final touches to the layout in word.

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 worked to create the tidyverse libray. This means that we are working in an open source library, and there are several different “authors” associated with the library. Working with open sources such as these would not usually be considered trustworthy according to most typical methods used by historians (henvisning til Erslev og noget SG), 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 open source, when using digital methods, can act as a sort 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()  
## )

## 4. Data Acquisition and Processing

**List and cite all sources of data used in this paper** **Details of data extraction, filtering and preparation. Attach processing scripts where relevant.**

The analysis of the dataset will take its stand in the visualisation of the data. 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/). This is a website run by historians from Aarhus Universitet, and is considered a trustworthy source for historical information.

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.

## 5. Implementation and Empirical Results

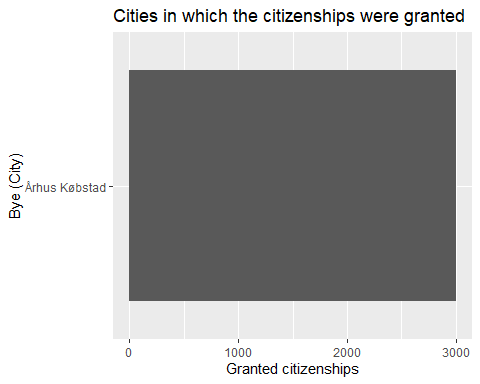
**Implementation details, or the full script demonstrating and documenting all major functions and decisions behind them** **Empirical Results (product of your script ~slides, map, outline, timeline…)**

## Visualization

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).

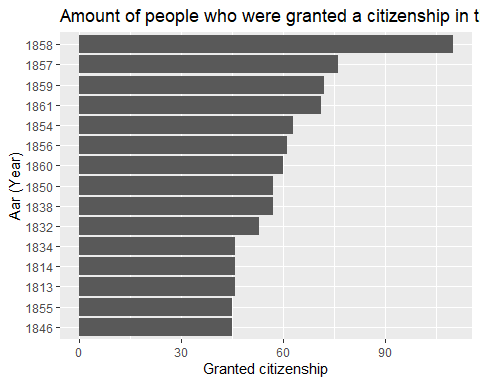
Citizenships %>%  
 count(bye, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(bye = reorder(bye, n)) %>%  
 ggplot(aes(x = bye, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "Bye (City)",  
 y = "Granted citizenships",  
 title = "Cities in which the citizenships were granted")

## 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.

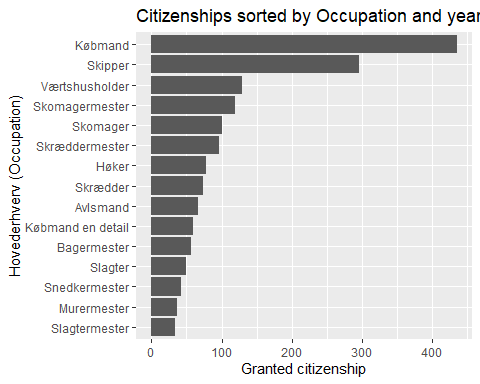
Citizenships %>%  
 count(aar, sort = TRUE) %>%  
 top\_n(15) %>%  
 mutate(aar = reorder(aar, n)) %>%  
 ggplot(aes(x = aar, y = n)) +  
 geom\_col() +  
 xlab(NULL) +  
 coord\_flip() +  
 labs(x = "Aar (Year)",  
 y = "Granted citizenship",  
 title = "Amount of people who were granted a citizenship 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 1857 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 order to verify the thesis that the spike in citizenships from 1857 to 1859 was a product of the new law on freedom of trade, I took a look at the occupations of these new citizens.

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 (Occupation)",  
 y = "Granted citizenship",  
 title = "Citizenships sorted by Occupation and year")

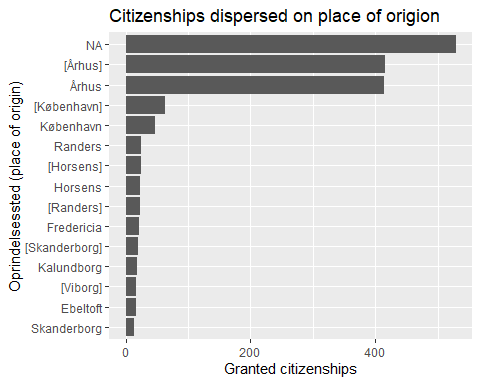
## Selecting by n

 After 1857, the citizenships were granted primarily to merchants and sailors. As a merchant, the limitations on the revenue were more fluid than in the more classical occupations such as carpenter, butcher or baker. This means that the chances of having a substantial (and maybe even growing) income when you were working as a merchant was significantly higher than the chances of the same if you were working with a handcraft such as carpentry. This could be the explanation as to why so many merchants were granted citizenships; they might not have been merchants by birthright, but with the new law, they may have decided to take their chances in this profession as opposed to their original line of profession.

To see if I could deduce any more relevant information from this dataset, I then set out to see where the new citizens had migrated from.

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 (place of origin)",  
 y = "Granted citizenships",  
 title = "Citizenships dispersed on place of origion")

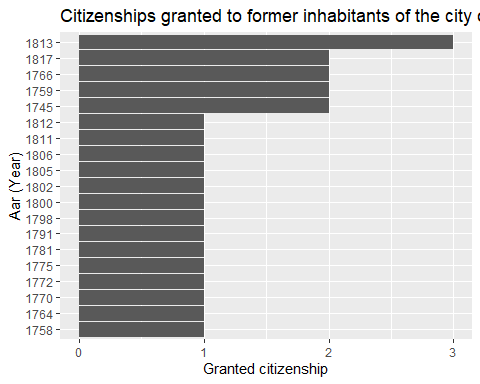
## Selecting by n

   
As this graph implies, the origin of the citizens were for a large part left unanswered (NA). It is not possible to say with any certainty why so many have been left unanswered in this category. It could be an expression of indifference when the data was first written down, but it could just as well have been an unwillingness from the aplicants for citizenship to announce their place of origion.

Lastly, I decided to see if there was a similar spike in citizenships granted as in graph 2, if I isolated a specific place of origin, in this case the city of Randers.

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 (Year)",  
 y = "Granted citizenship",  
 title = "Citizenships granted to former inhabitants of the city of Randers")

## Selecting by n



The city of Randers, which was chosen randomely from the list of places of origin, did not have a similar spike around the years 1857-1859. In fact, it did not have any registered citizens after 1817, where 2 people applied to get their citizenships transfered to Aarhus. Again, it is not possible to say with any certainty why there are no informations about citizens from Randers after 1817.

## 6. Critical evaluation

**Comparison with the state-of-the-art software if any (kindly cite relevant work, scripts, etc.)** **Evaluation of the learning process, time on task, vis-à-vis the product.**

As discussed in section 3 with the title Software Framework, the scripts from the tidyverse - and RStudio in general, are open source. This means that it is not traceable back to a single author. Using other historiographic methods, this could pose a threat to the credibility of the source, but as stated earlier, the open source enhances the chances of catching the flaws in any given script. On this basis, it is concluded that the RStudio and tidyverse are both trustworthy scripts when working with the digital methods in a historical research.

The digital methods are highly usable in a case like this, to ensure a quick overview of the data, without having to scroll through the entire list of citizens.

Don’t know anything about state-of-the-art software. Maybe a comparison to the original script by MAx. My script understandably looks very much like his. Had I had a deeper understanding of the tools at hand, I imagine I could have had a variety of different types of visualizations. A few of my considerations regarding the extraction of information from the dataset through the visualization, were dropped so as not to exeed my own limits of understanding.

## 7. Conclusions

**Set out the conclusion of this original software publication**

## References