Digital Methods: A comparison of Marital Statuses from 1801 and 1845

Author: Frederikke Bakkestrøm

RStudio - R i386 3.6.1

Table of Contents

[1.0 Abstract. 2](#_Toc29306269)

[2.0 Keywords: 2](#_Toc29306270)

[3.0 Introduction 2](#_Toc29306271)

[5.0 Problems and Background 3](#_Toc29306272)

[6.0 Software Framework 3](#_Toc29306273)

[Github, gitplugg???? deodkeopwajq 3](#_Toc29306274)

[7.0 Data Acquisition and Processing 3](#_Toc29306275)

[8.0 Implementation and Empirical Results 4](#_Toc29306276)

[8.1 Marital Statuses sorted by gender 4](#_Toc29306277)

[8.2 Marital Statuses sorted by geography 6](#_Toc29306278)

[8.3 Marital Statuses sorted by age 8](#_Toc29306279)

[8.3 Marital Statuses sorted by age 10](#_Toc29306280)

[9.0 Results 11](#_Toc29306281)

[10.0 Critical evaluation 11](#_Toc29306282)

[11.0 Conclusions 12](#_Toc29306283)

[12.0 Leaning Journal 13](#_Toc29306284)

[Assignment: Stopwords from Regex to Voyant 13](#_Toc29306285)

[Working with R Studio: 14](#_Toc29306286)

[Thoughts/intention 14](#_Toc29306287)

[13.0 References 15](#_Toc29306288)

[B- Required Metadata FÅ HJÆLP 15](#_Toc29306289)

[B1 Current executable software version 15](#_Toc29306290)

[Table 1 – Software metadata 16](#_Toc29306291)

# 1.0 Abstract.

Based on censuses from Aarhus Stadsarkiv, the focus of this assignment will be a comparison between censuses from Aarhus Stadsarkiv of Marital Statuses from the year 1801 to 1845. These censuses contain data information about the citizens marital statuses, which leaves three different possibilities; married, unmarried and widowers. With these datasets, the project will look into possible similarities and differences in marital status, both in total numbers of unmarried and married, in gender, age and also geography, taking to different counties in Jutland. Most importantly the assignment will investigate whether these numbers have changed during period of 44 years. In order to do this, the assignment have used the () Rstudio. By plotting several codes into Rstudio, the datasets gathered from Aarhus Stadsarkiv was turned into better visualised tables. The results of the comparison between the two datasets, censuses from 1801 and 1815, was tough not revealing anything (). It revealed no big difference in the regards of gender, though it did reveal a change in the number of unmarried/married in both genders from the censuses made in 1801 and the censuses in 1845.

# 2.0 Keywords:

Married; unmarried; widows; Randers; Aarhus; ; Hvad mere????

# 3.0 Introduction

In modern days statistics reveals that less people choose to get married, and it is easy to be caught in a sentimental “back in the old days” state of mind, fantasising about a time, when everyone strived for marriage and did so in an early age. It could be interesting to find out, what was the actual case in the marital (state) in Denmark in earlier days, as to how many people were unmarried or married, and what gender and age was the most common revealed in these statistics. Furthermore, how these numbers have evolved throughout the years. Also if geography played a part.

The world we know is undeniably becoming a more digitalized one and is undouble going to effect the work of historians. The motivation behind developing this script was to show the possibilities that R Studio can provide. With Rstudio it is possible to make censuses from Aarhus Stadsarkiv, which as a historian is the kind of da quite often, from raw data into nicely visualized data. I found this specific script to be great to quickly pull out specific data and make it visually understandable, especially to illustrate differences and similarities. For a historian in this specific case, the task could be writing a book about marital status in the 19th century, where one could pull out the relevant information from raw datasets themselves, without hiring in specialists. It is important for historians to know these simple digital skills in a world that becomes more digitalized every day and this assignment seemed to be a good way to start.

# 5.0 Problems and Background

Beskriv eventuelle problemer og baggrund for din opgave

# 6.0 Software Framework

3.1 Software Architecture/Prerequisites

Give a short overview of the overall software architecture, dependencies and prerequisitesGithub, gitplugg???? deodkeopwajq

# 7.0 Data Acquisition and Processing

**Introducing dplyr, tidyverse and Starting with DATA in RStudio???? Få Andreas til at hjælpe**

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelseFÅ INDSAT DETTE FRA WORD!

# 

# 8.0 Implementation and Empirical Results

## 8.1 Marital Statuses sorted by gender

#### 8.1.1 Tables based on censuses from 1801 from Aarhus Stadsarkiv

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

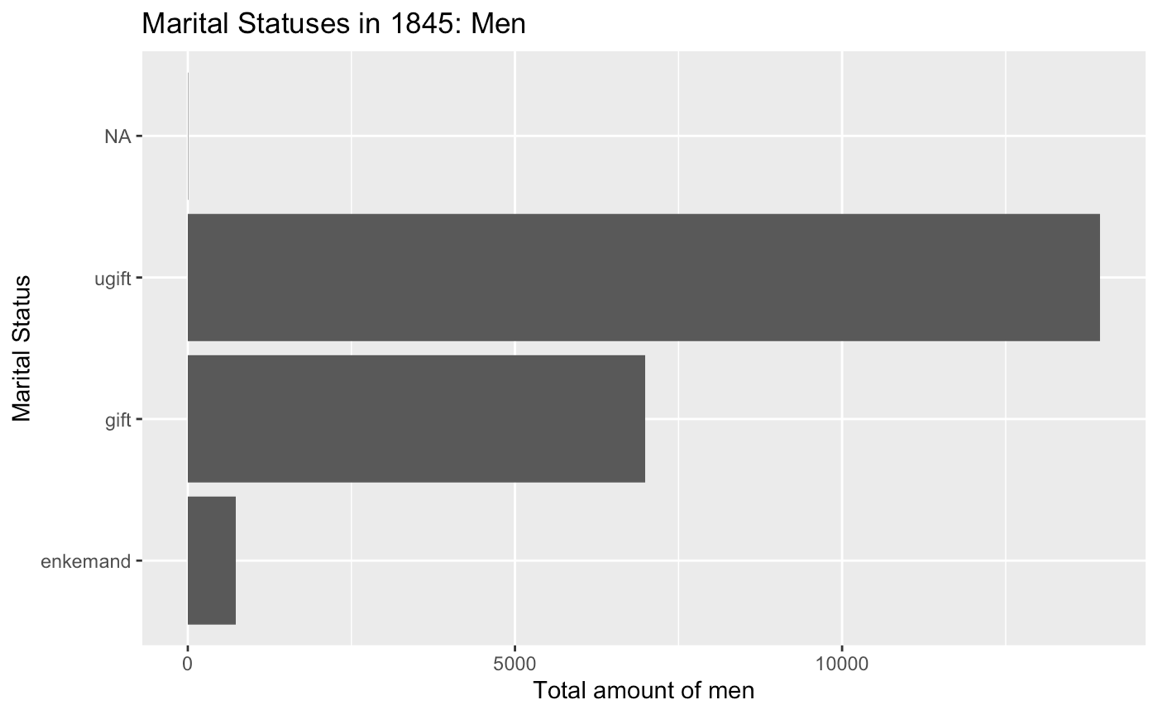
Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

#### 8.1.2 Tables based on censuses from 1845 from Aarhus Stadsarkiv

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse



## 8.2 Marital Statuses sorted by geography

8.2.1 Tables based on censuses from 1801 from Aarhus Stadsarkiv Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

8.2.2 Tables based on censuses from 1845 from Aarhus Stadsarkiv

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse

Et billede, der indeholder skærmbillede

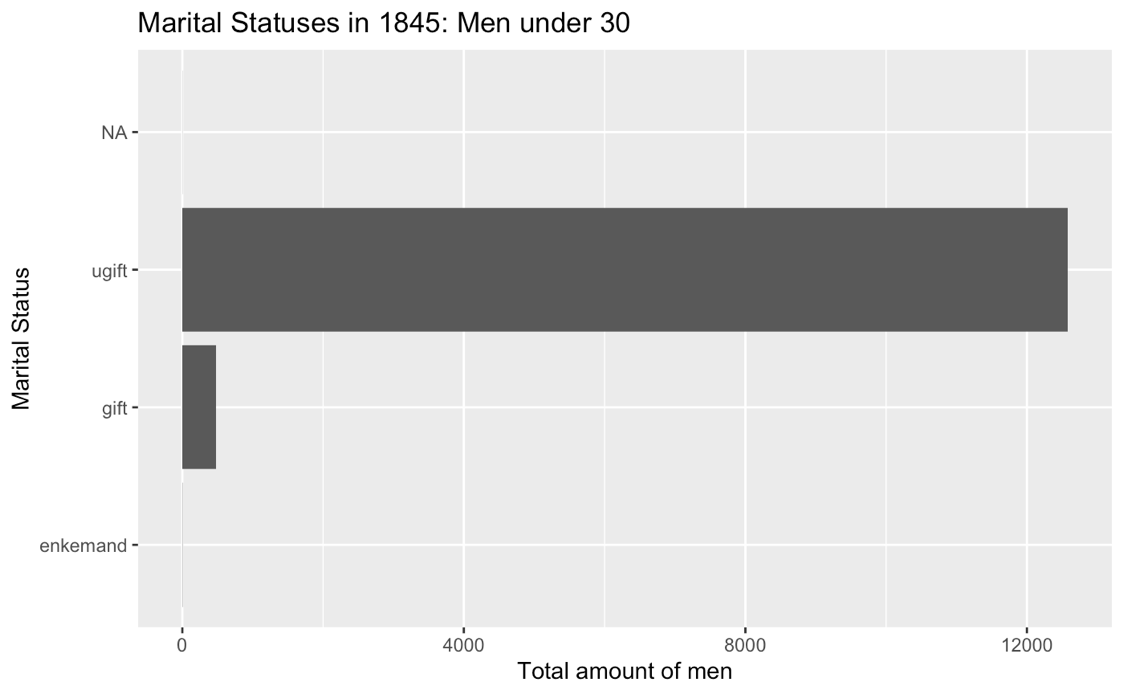
Automatisk genereret beskrivelse

## 8.3 Marital Statuses sorted by age

#### 8.3.1 Tables based on censuses from 1845 from Aarhus Stadsarkiv

Et billede, der indeholder skærmbillede

Automatisk genereret beskrivelse



## 8.3 Marital Statuses sorted by age

#### 8.3.1 Tables based on censuses from 1845 from Aarhus Stadsarkiv

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…)

# 9.0 Results

Generally the results came out looking quite alike. Based on the censuses from Aarhus Stadsarkiv, the tables that visualizes marital statuses in the years 1801 and 1845 does not show significant changes in the gap between unmarried and married. Though the tables reveal that in 1801 under 12.500 women were unmarried and around 8000 were married, whereas in 1845 more than 12.500 women were unmarried and 7000 married. This shows, that fewer women in this period of time got married. In both genders the number of married are around 8000 in 1801. And as for the women also fewer men got married as revealed in the censuses from 1845. The numbers fell from around 8000 in 1801 to a shade under 7000 in 1845. Furthermore, the tables regarding gender shows that there were more unmarried men than women in 1801. Whereas under 12.500 women were unmarried there were approximately more than 13.000 men.

Regarding both gender and age, the tables reveal a difference ……..

As goes for the comparison based on geography, comparing two different cities, Randers and Aarhus, (kom med et par ord mere) there is not much to say, but it did help explore my abilities to work in Rstudio further.

Age:

# 10.0 Critical evaluation

As far as this project goes there was also some struggles a long the way. Firstly, I expierienced issues with managing danish words in the datasets, as for instance the letter “Å”. Though Rstudio did eventually for some reason that I do not know, accept it. And this exact problem reflects perfectly, how digital methods for beginners are hard since you do not yet understand every problem that can appear and thereby know how to solve it. The difference in language is also revealed in the tables, that I have not been able to make (ens).

In general, the two datasets used in this project seemed quite net though another possible dataset which I did not include regarding censuses from 1787, one of the listed ages was 764. If I had chosen to work with that dataset, I should have left that one individual out of the data processing to avoid a wrong result.

A factor to be aware of in this specific project is the difference in unmarried men and women in 1801, that was visualised in the tables, since there might in general have been more men, and therefore the statistics(eller ??) might in fact not be so far from each other at all.

Even though the project didn’t reveal the most eye lifting results, all of these different illustrations made became a great way to learn, how you can work with different datasets in Rstudio. For example I leaned to quickly pull out the relevant information about age, gender, (demografi) using different filters in the codes. Though I did not get to figure out how to, if it is even possible in Rstudio, to make the differences more clearly by bringing two different () in the same table. That would inquisitional add more to a project like this project though it must be a job for the future investigation of digital methods.

The progress of this projects reflects exactly why historians as humanitarians, may not yet be masters of digital methods. It was a really difficult progress with a lot of needed troubleshooting and professional guidance. Mastering digital methods turned out to be comparable to learning a new language and therefore the learning progress is clearly not a not finished chapter with this assignment being done. It will take much more hands-on work in the future. Though it did indeed reveal the effectiveness and positive outcomes that even easier (programmer?) like Rstudio can give anyone who wishes to illustrate data. Another great help in learning to work in Rstudio was writing a learning journal so that relevant information in every class work through and in my own ,,,, was kept, so my learning journal came in handy when I was to make this project. This specific part of the learning journal is included in the end of this project.

# 11.0 Conclusions

Plotting in datasets with censuses from Aarhus Stadsarkiv from the years 1801 and 1845 into RStudio the project could better draw out relevant information about marital statuses in Denmark. The results was perhaps not (), but it did reveal a small differences between male and female marital statuses and did also show a change in both genders martial statuses in the 44 years, from the censuses made in 1801 and the ones made in 1845. In this timespan the number of particularly male marriages fell, from around kdoekdoek to dkoekdoke in 1845. The number of unmarried (steg?) also in both genders. The purpose of this assignment was as (fremlagt) in the introduction; to investigate whether more people got married in earlier days, and giving the falling in number of marriages revealed in the tables made in Rstudio, it can be concluded, that in 1801 more people did get married than they did in 1845, 44 years later. If the view on the gap between married and unmarried should have been more clear data from 1801 and up to today should all have been included in the data processing. But based on the datasets a change has definitely accured.

The age…..

# 12.0 Leaning Journal

These are relevant (udpluk) from the learning journal with a special focus on learning simpler tasks as converting stopwords from Voyant to R and vise versa, but also in the progress of leaning how to work with R Studio, that I used in this assignment.

## Assignment: Stopwords from Regex to Voyant

#### Thoughts/intention

The point of this specific assignment was to test my ability to convert the stop-words, so that it would be usable in Voyant and Regex respectively. Therefor, I found out the differences of what kind of text, that will be accepted in the programs of Voyant and Regex. Voyant will for instance not accept quotation signs (”) nor will it accept commas (,).

#### Action

In working on this assignment, firstly I sat down together with some of my class mates and brainstormed – to make sure we had all understood the assignment correctly. Then I initiated the work by plotting in all the Regex’ stopwords into the ’TextingString’ box in Regex.

What my goal with these stopwords was, was to make them stand in line underneath one another without quotation signs or commas, so that the texts would be compatible with Voyant. In order to do so, I removed the signs in the text, which is not compatible with Voyant by typing ”, ” into the ’Regular Expression’ box. Thereafter, I had to remove the signs, which I just marked throughout the entire text and to do so, I had to susbtitute these signs with a new line. The regular expression for a new line in Regex is – so I scrolled down to the ’Substitution‘ box. The results I got turned out exactly what I wanted from typing the regular expression commands into Regex: the text without signs and standing in line. Though Regex have clearly not been able to remove the first and last of the quotation signs, and for a simple solution, I went back to the editing box, and manually remove them.

#### Final results

The final results made the words compatible with Voyant. This has been a great first hands on with digital history.

## Working with R Studio:

### Thoughts/intention

After some of the first experiences with digital data programmes, I scheduled a trouble shooting meeting with Max Odsbjerg, and that gave me the idea, as to what programme I wanted to use in my final project. I played around with datasets on marriage statistics from Aarhus Stadsarkiv in RStudio

#### Action

Firstly I got the statistics from Aarhus Stadsarkiv, but when loading them I was told the package was too bit for Rstudio, and that I could only get the data "raw". But using the URL in the Rstudio programme with this code lpdeldepldpelde raw data could be converting into nicely looking visuals.

Now the dataset was () to appear more visually easy in a diagram using this code…. Furthermore to make this visualisation, I chose a specific danish parish, and inserted it in the filter box. Firstly I typed "sogn" and then I typed in the parish "Hørning" which makes Rstudio select the data in the column of the parish of *Hørning: "filter(amt == “Skanderborg”).*

I also typed in 'aar' in the count box, to let Rstudio futher know how to sort the data. The X in the diagram should now be 'aar'(years) and the y should be the number of marriages. Carrying on with the data processing, I tried making a diagram on this dataset with counties instead of parishes. I chose the county of 'Skanderborg' and insertet both these informations in the filter box exactly the way I did before: *filter(amt == “Skanderborg”).*

Trying yet one more time to play around with this filter box, I tried to make a diagram based on wedding dates. Just as an example I chose the date 24/09. What I had to be attentive to is the way the date is written in the original data set. 24/09 is written "24de Septbr." and therefor that is the way I had to write it too in the filter box, appearing like this: *filter (vielsesdato == "24de Septbr.")*

**Introducing dplyr, tidyverse and Starting with DATA in RStudio???? Få Andreas til at hjælpe**

#### Final thoughts

This first hand experience with RStudio made me certain that Rstudio should be the program to use in my final project.

# 13.0 References

* https://github.com/aarhusstadsarkiv/datasets
* https://github.com/aarhusstadsarkiv/datasets/tree/master/censuses
* https://rstudio.com/
* https://cran.r-project.org/web/packages/tidyverse/index.html
* https://r4ds.had.co.nz/index.html#
* https://www.tidyverse.org/
* <https://raw.githubusercontent.com/aarhusstadsarkiv/datasets/master/censuses/1801/census-1801-normalized.csv>
* <https://raw.githubusercontent.com/aarhusstadsarkiv/datasets/master/censuses/1845/census-1845-normalized.csv>

# 

# B- Required Metadata FÅ HJÆLP

# B1 Current executable software version

Ancillary data table required for sub version of the executable software: (x.1, x.2 etc.) kindly replace examples in right column with the correct information about your executables, and leave the left columns as they are

### Table 1 – Software metadata

|  |  |  |
| --- | --- | --- |
| Nr | (executable) Software metadata description | *Please fill in this column* |
| S1 | Current software version | RStudio - R i386 3.6.1 |
| S2 | Permanent link to executables of this version | *example : https://github.com/combogenomics/DuctApe/releases/tag/DuctApe-0.16.4* |
| S3 | Legal Software License | *List one of the approved licenses* |
| S4 | Computing platform / Operating System | *Mac OS Mojave, version: 10.14.6 (18G87)* |
| S5 | Installation requirements & dependencies | *??* |
| S6 | If available Link to user manual - if formally published include a reference to the publication in the reference list | *Example http://mozart.github.io/documentation/ or* |
| S6 | Support email for questions |  |