Abortion in Denmark since 1973

Authors: Mathilde Halberg Nielsen og Trine Thestrup Larsen

Keywords: Abort data, Visualizations, Contribution to a clearer debate, Rmarkdown

Table of Contents

Introduction (Trine and Mathilde)	2
Problem orientation and Background (Trine and Mathilde)	2
Software Framework (Trine and Mathilde)	4
Data Acquisition and Processing (Trine and Mathilde)	4
Visualizing the number of registered abortions since 1973 (everything coded together, Trine wrote this part here)	5
Visualizing the development in the number of women since 1973 to 2022 (everything coded together, Trine wrote this part here)	
Making a stacked plot visualizing in which weeks the abortions were performed between 1973 and 2022 (everything coded together, Mathilde wrote this part here)	. 15
Empirical Results (Trine and Mathilde)	.21
Lessons learned in general while working with this project (Trine and Mathilde)	. 25
Critical evaluation (Trine and Mathilde)	26
Conclusions (Trine and Mathilde)	.27
Acknowledgements (Trine and Mathilde)	28
References	28
Table 1 – Software metadata (Trine and Mathilde)	
Table 2 – Data metadata	31

Introduction (Trine and Mathilde)

"If a mother can kill her own child - what is left for me to kill you and you to kill me - there is nothing in between." Is a quote taken from the world-famous catholic nun Mother Teresa and the following originates from the famous nurse, sex educator and writer Margaret Sanger: "No woman can call herself free who does not control her own body". Those are just some of the heated quotes about the opinions covering abortion in the world and through time. Even though these quotes are old, the abortion debate is still just as heated today as it has been throughout time and is often divided into pro-life and pro-choice groups. The right to an abortion is different in each country around the world, but has been legalized in several of them, including Denmark in 1973². Even though there is an existing legislation in many countries, though with different rights and boundaries in each, it is still a much-debated topic that activates strong feelings for many people and which has been debated both religiously, politically, historically, ideologically and individually. In Denmark the debate has recently become relevant again and therefore, by working with data on the abortions rates in Denmark, we hope to contribute to a more positive and clearer debate based more on facts than assumptions.

Problem orientation and Background (Trine and Mathilde)

This year, 2023, is the 50th anniversary of free abortion in Denmark. In 1973, Denmark was one of the first west European countries to legalize free abortion³. The access to the free abortion was the final result of several political and social developments, which since the 1930's, where 'Svangerskabslovgivningen' and 'Mødrehjælpen' were created, had opened for new possibilities to get an abortion under certain circumstances and had created an increasingly heated debate about this right⁴. However, it was not only a result of a 40-year long debate, but came in the wake of many centuries where illegal abortions were the only option if you wanted to get an abortion. This often included many risks for the woman and

¹Goodreads (2023) "Abortion quotes" https://www.goodreads.com/quotes/tag/abortion

²Sørensen, Anne (3.oktober 2017) "Fri abort, 1973-"https://danmarkshistorien.dk/vis/materiale/fri-abort

³ Sørensen, Anne (3.oktober 2017) "Fri abort, 1973-"https://danmarkshistorien.dk/vis/materiale/fri-abort

⁴ Hytting, Emmelie Sakina (27.februar 2010) "Tidslinje: Abortens udvikling i Danmark" (https://www.etik.dk/etik/tidslinje-abortens-udvikling-i-danmark)

for the people involved in the affair. It is estimated that in the decades preceding 1973, there were between 15.000-25.000 illegal abortions in Denmark each year⁵

Throughout time, the opinion on abortion has been affected by both secular and ecclesiastical rules, punished with death penalties until 1866, prison sentences from 1866 until the 1930's and general stigmatization and taboo from society, a taboo which still more or less exists in Denmark⁶⁷.

However, the anniversary also creates a fertile ground to debate the legislation surrounding abortion and this same year the ethical counsel, 'Etisk Råd', in Denmark, a counsel consisting of 17 members appointed for 3 years each and who are both professional and lay persons discussing ethical questions in society, discussed whether the limit for free access to provoked abortions should be extended based on the change in fetal diagnostics which was introduced in 2004⁸⁹. At this point it is only possible to get a provoked abortion within the first 12 weeks of the pregnancy, except when it concerns special cases. In September 2023, 'Etisk Råd' decided to recommend an abortion limit at 18 weeks, the same limit that persists in Sweden¹⁰¹¹. Today, there is a certain number of Danish women every year who travel to Sweden and England to have an abortion after 12 weeks. In England the right to a free abortion expires after 24 weeks¹²¹³.

The debate surrounding abortion is often very emotional and can be based on assumptions. We hope that this digital project, which will consist of visualizations, will contribute to pushing the debates surrounding the subject in a more fact-based direction. Thereby we hope to make the data more manageable for the majority and to contribute to a more fact-based discussion.

⁵Sørensen, Anne (3.oktober 2017) "Fri abort, 1973-"https://danmarkshistorien.dk/vis/materiale/fri-abort

⁶ Sørensen, Anne (3.oktober 2017) "Fri abort, 1973-"https://danmarkshistorien.dk/vis/materiale/fri-abort

⁷ Hytting, Emmelie Sakina (27.februar 2010) "Tidslinje: Abortens udvikling i Danmark" https://www.etik.dk/etik/tidslinje-abortens-udvikling-i-danmark

⁸Folketinget.dk (2023) "Det Etiske Råds udtalelse om en eventuel ændring af abortgrænsen" https://www.ft.dk/samling/20061/almdel/uer/bilag/12/359392.pdf

⁹ Nationalt Center for Etik (2023) "Om det Etiske Råd" https://nationaltcenterforetik.dk/raad-og-komiteer/detetiske-raad/om-det-etiske-raad/om-det-etiske-raad/
https://nationaltcenterforetik.dk/raad-og-komiteer/det-etiske-raad/om-det-etiske-raad/
https://nationaltcenterforetik.dk/raad-og-komiteer/det-etiske-raad/
<a href="https://nationaltcenterforetik.dk/raad-og-komiteer

¹⁰ Nielsen, Peter Ingemann (26.september 2023) "Flertal i Etisk Råd anbefaler at hæve abortgrænsen" https://www.altinget.dk/artikel/flertal-i-etisk-raad-anbefaler-at-haeve-abortgrænsen

¹¹Heissel, Anders (26.september 2023) "Et flertal i Etisk Råd anbefaler at hæve abortgrænsen til 18 uger". DR. https://www.dr.dk/nyheder/indland/et-flertal-i-det-etiske-raad-anbefaler-haeve-abortgrænsen-til-18-uger

¹² Birk, Christian (16.april 2018) "Danske kvinder tager til udlandet for abort" https://www.kristeligt-dagblad.dk/danmark/danske-kvinder-tager-til-udlandet-abort

¹³Heissel, Anders (25.september, 2023) "Svensk jordemoder hjælper danske kvinder med sen abort : Jeg er glad for at kunne hjælpe" https://www.dr.dk/nyheder/indland/svensk-jordemoder-hjaelper-danske-kvinder-med-sen-abort-jeg-er-glad-kunne-hjaelpe

Software Framework (Trine and Mathilde)

We wrote the code for this project on a 3-year-old Macbook Pro from 2020 version 13.6 macOS Ventura. We have used the desktop version of RStudio, version 2023.09.0+463, and more specifically R Markdown as our digital platform for coding, tidying our data and creating visualizations.

Data Acquisition and Processing (Trine and Mathilde)

We have acquired most of our data from the register of *Sundhedsdatastyrelsen* concerning abortions¹⁴. They have published a dataset on their platform esundhed.dk which includes research on four indicators: 1) The number of abortions from 1973-2022, 2) The method used, 3) The number of induced abortions by length of pregnancy and 4) Number of induced abortions by age groups. We decided to focus on and work with number 1 and 3 after the judgment that those would be the most relevant for our project and our focus¹⁵.

At first sight we thought that the data was ready for immediate use for us to work with in Rmarkdown and for us to use when creating our visuals, which was our main goal, but little did we know. What seemed like tidy data for the human eye turned out to be rather messy and difficult to read for our coding program, once we tried to work with it in Rmarkdown.

We consulted the book *R for Data Science*¹⁶, focusing on the chapter named "Data Tidying", which states that getting data in the right order takes some work, but that it pays off in the long term and gives more time to focus on the actual project. For us, creating visualizations.

We have also used data from *Danmarks Statistik* concerning the development in the number of women in Denmark from 1973 to 2022¹⁷, to cover the same years reported in the data from *Sundhedsdatastyrelsen* on the number of abortions. We discovered the same issue with this dataset as with the data on abortions; that it was published in a format that is more

¹⁴ Sundhedsdatastyrelsen and Esundhed (2023) 'Register over Legalt Provokerede Aborter (ABR) https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Provokerede-aborter

¹⁵ Sundhedsdatastyrelsen and Esundhed (2023) 'Register over Legalt Provokerede Aborter (ABR) https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Provokerede-aborter → look in the excel file ¹⁶Wickham, Hadley et al. (2023) "R for Data Science (2e)". Chapter 5. Data Tidying. O'Reilly. https://r4ds.hadley.nz/

¹⁷ Danmarks Statistik (2023): "BEFOLK2, Befolkningen 2.januar efter køn og alder (1901-2023) click on: 'Kvinder', 'Alder i alt' and '1973-2022) https://www.statistikbanken.dk/20021?fbclid=IwAR0OiMFvfwD8sPY4oXTJPqPIgAn7uMn5lULCrzf43wi5vBxYMwlR-F9HiFY

Studienummer: 202008537

Studienummer: 202008125

accessible for humans than R. Therefore, our task with this was also to make it tidy, before we would be able to create the visualizations, with which we wanted to illustrate our topic. We discovered, also by consulting the chapter "Data Tidying" in *R for Data Science* ¹⁸, that Rmarkdown is more inclined to work with so-called long data and that the function pivot is essential for this. Since all of our data was in a wide format, our first task was therefore to transform it, and while doing this we realized that we needed to discover new functions.

Functions which we will elaborate in the following section.

We also searched for the permissions to use both datasets for our analyses and found them on their respective homepages. They write that it is permitted to use their data, as long as they are referenced in our project, which we have done several times ¹⁹²⁰.

Visualizing the number of registered abortions since 1973 (everything coded together, Trine wrote this part here)

Our starting point was the data concerning the number of registered abortions between the years of 1973 and 2022, which we wanted to visualize in a readable and transparent graph to show how the abortion trend has developed since the abortion became a free option in Denmark in 1973. It is also possible to follow all the steps in the Rmarkdown file uploaded in each of our GitHubs.

We started by installing and loading the tidyverse package, which included many of the basic packages we needed. We also fingered that to load our excel file with the data, we needed the package named "readxl", which we also loaded from the library:

library(tidyverse)
library(readxl)

Afterwards we made Rmarkdown load our excel file from our data folder containing our data from Sundhedsdatsstyrelsen by using the following code:

Wickham, Hadley et al. (2023) "R for Data Science (2e)". Chapter 16 Factor. https://r4ds.hadley.nz/

¹⁹ E-sundhed (2023) "Anvendelsesvilkår" https://www.esundhed.dk/Anvendelsesvilkaar

²⁰ Danmarks Statistik (2023) "Kildeangivelse og brug af Danmarks Statistiks logo" https://www.dst.dk/da/Statistik/brug-statistikken/kildeangivelse

Studienummer: 202008537

Studienummer: 202008125

```
Abortdata <- read_excel("data/Abortdata.xlsx", sheet = "Tabel 1 - Antal", col_names = FALSE, skip = 1)
```

We created the code above by clicking on the "Import Dataset button" in the head bar in Rmarkdown and selected the option named "From Excel". Then we passed the Excel file concerning numbers of abortions from the years 1973 to 2022. Our data from Sundhedsdatastyrelsen consists of several tables and we choose number one. We also had the opportunity to skip the first row of our table, consisting of nothing but text, which we did. As a result, we got a "Code Preview", which we copied and ran in our code field.

We could then discover that our data was constructed in a non-readable format for R, and that we had to tidy the format.

First, we located some unsuitable rows, which weren't relevant for our case and which we wanted to get rid of. For that purpose, we used the function called "slice" from the "dplyr" package. However we started by assigning that we would like to work with our "Abortdata" and afterwards we used a pipe-operation: "%>%", an operation which sends the data further in the calculation. Then we used the slice function and indicated which rows we wanted to cut off and assigned the result to our data using the assign operator: "->".

```
Abortdata %>% slice(2:3) -> Abortdata
```

Afterwards we wanted the first row of our table to become row names and for that purpose we found a package called "janitor", which includes a function named "row to names".

Abortdata %>% row to names(row number = 1) -> Abortdata

We read in R for data science²¹, that it is important and a basic rule for tidy data, that all the variables in a table are in columns, that those columns are in rows and that the characters should be the same data type, we wanted to comply with these rules in the following.

First, we transformed all the characters to numeric using the "mutate_if" function from the "dplyr" package.

```
Abortdata %>% mutate if(is.character,as.numeric) -> Abortdata
```

Afterwards we transformed the format of the data from a wide format to a long format, we turned the table around so to speak. To do so we used the "pivot_longer" function from the "tidyverse package". We selected the columns in the data and assigned them to separate columns with the names "Year" and "Abortions", which would later become the values on our x and y axis in the ggplot.

After we rearranged the table, we ended up with a column named "År", which we wanted to get rid of. For that purpose, we also used the "select" function from the "dplyr" package to minus the column with "År".

```
Abortdata %>%

pivot_longer(cols = 2:51,

names_to = "Year",

values_to = "Abortions") %>%

select(-År) -> Abortdata
```

We had then tidied our data and could begin creating the visualization. We decided on a geom point visualization in color, since we thought it to be most fitted for the job of creating a simple visualization. We used our primary tool for visualizations; a package called "ggplot2". At first, we assigned the "year-column" to the x-axis and the "abortion-colum" to the y-axis. Afterwards we choseed the "geom_point" function from the "ggplot2" package and asked for brown dots in our visualization.

²¹ Wickham, Hadley et al. (2023) "R for Data Science (2e)". Chapter 5 Data Tidying. https://r4ds.hadley.nz/

We discovered that the different years on our x-axis were unreadable at first, since all the numbers was cluttered and stacked on top of each other. We then found a function in the "ggplot2" package named "theme()", which we used to turn the years 90 degrees, and thereby made them possible to read. We also changed the title of the visualization by using the funktion called "labs()".

```
Abortdata %>%

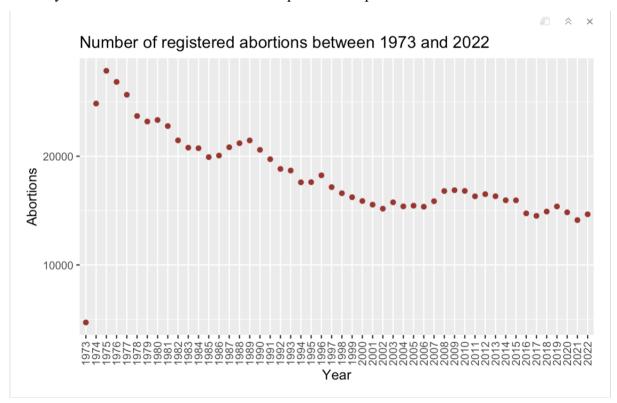
ggplot(aes(x = Year, y = Abortions))+

geom_point(color = "brown")+

theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+

labs(title = "Number of registered abortions between 1973 and 2022")
```

Below you can see our visualization at this point of the process.



We weren't visually satisfied with the very strong colors that R could provide, and we wanted to use colors from the same palette for all our visualizations to create a cohesion in our

assignment. Therefore, we tried to make new colors for the graph by loading a package from the filmic universe of Wes Anderson named "wesanderson"²².

```
install.packages("wesanderson")
library(wesanderson)
```

In the code below we used the fourth color from the color scale named "GrandBudapest1"

```
Abortdata %>%

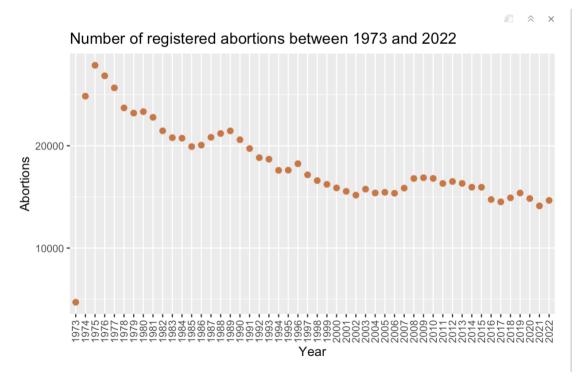
ggplot(aes(x = Year, y = Abortions))+

geom_point(color = wes_palette("GrandBudapest1")[4] +

theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+

labs(title = "Number of registered abortions between 1973 and 2022")
```

Below you can see the final result for our geom point graph:



²² Ram, Karthik (no exact date) "Wes Anderson Palettes". Github. https://github.com/karthik/wesanderson (found 30 October 2023)

Visualizing the development in the number of women since 1973 to 2022 (everything coded together, Trine wrote this part here)

We wanted to support our geom plot graph concerning the number of abortions between the years 1973 and 2022, and therefore we decided to create another geom_point plot, regarding the population of women in Denmark at the same time span 1973 to 2022. Thereby we had the possibility to compare the two graphs and show if there was a cohesion between the number of women and the number of abortions in Denmark. An interpretation we will return to later in the paper. To do so we had to work with a new dataset from "Danmarks Statistik" in an excel format, which we implemented in Rmarkdown the same way as we did with the previous dataset.

```
library(readxl)
Antal_kvinder <- read_excel("data/Antal_kvinder.xlsx",
    sheet = "BEFOLK2", col_names = FALSE,
    skip = 1)</pre>
```

We noticed that the dataset wasn't fit for the making of visualizations in R and that we had to perform a data transformation once again and tidy our data.

First, we cutted of some unwanted rows using the previously mentioned "slice" function.

```
Antal_kvinder %>% slice(2:3) -> Antal_kvinder
```

Afterwards we wanted the first row to be the column names and used the "janitor" package and the once before mentioned "row_to_names()" function.

```
Antal kvinder %>% row to names(row number = 1) -> Antal kvinder
```

²³ Danmarks Statistik (2023): "BEFOLK2, Befolkningen 2.januar efter køn og alder (1901-2023) click on: 'Kvinder', 'Alder i alt' and '1973-2022) https://www.statistikbanken.dk/20021?fbclid=IwAR00iMFvfwD8sPY4oXTJPqPIgAn7uMn5lULCrzf43wi5vB xYMwlR-F9HiFY

As mentioned before R is most likely to work with data in a long format, and since our data was in a wide format we had to transform it using the "pivot_longer()" function. Afterwards we selected the columns in the data and assigned them to separate columns with the names "Year" and "Women" which would later be the values on our x and y axis in the ggplot. We also used setNames(paste0("a", names(.))) since we wanted to rename the data in the dataframe.

```
Antal_kvinder %>% setNames(paste0("a", names(.))) %>%

pivot_longer(cols = a1973:a2022,

names_to = "Year",

values to = "Women") -> Antal kvinder
```

We were met with an error and therefore we checked what our data consisted of with the following function:

```
sum(!complete.cases(Antal kvinder))
```

It was as we wanted, and we continued.

Somehow all of the years in our table also has an "a" in front of it, which we would like to remove using the function called "mutate()" from "dplyr".

```
Antal_kvinder %>%

mutate(Year = sub("a", "", Year)) -> Antal_kvinder
```

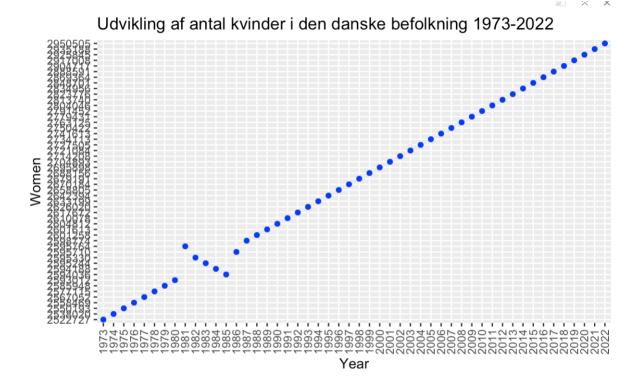
Afterwards we were done transforming our data. It was then tidy enough to be used in our visualization. We did it the same way as before by using "ggplot2", creating a "geom_point" graph and turning the years on the x-axis 90 degrees by the "theme()" function. However, this time we placed the number of women on the y-axis and made the dots blue as a start. We also named the graph "Udvikling af antal kvinder i den danske befolkning 1973-2022".

```
Antal_kvinder %>%

ggplot(aes(x = Year, y = Women))+
```

```
geom_point(color = "blue")+
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+
labs(title = "Udvikling af antal kvinder i den danske befolkning 1973-2022")
```

Below you can see the result of our visualization at this point



Afterwards we wanted to change our visualization and make the y-axis more readable and easier to overview. We wanted to sum the numbers and make it show every tenth. To do so we found a function named "scales()" from the "ggplot2" package.

library(scales)

After that, we wanted to convert and change the datatype for our column of women to numbers, to make sure everything was the same datatype, to be able to use the "scale" function again.

Antal kvinder\$Women <- as.numeric(Antal kvinder\$Women)

We then tried to make our visualization again, with a new name in english and used the scales function. We titled the graph as follows: "Development in number of women in the danish population from 1973-2022".

```
Antal_kvinder %>%

ggplot(aes(x = Year, y = Women))+

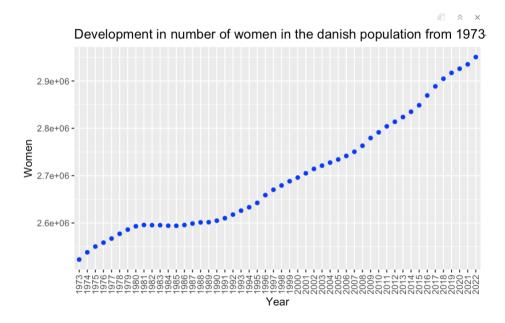
geom_point(color = "blue")+

theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+

labs(title = "Development in number of women in the danish population from 1973-2022")+

scale_y_continuous(labels = label_scientific())
```

This is the result so far



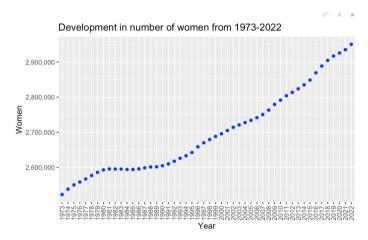
We now got what we wanted, but we would like to remove the "e + 06" from the numbers on the y-axis to make it more accessible and readable as mentioned above. "e + 06" is a mathematical term which expresses that the number comes in the power of 10. "e" stands for the exponent and "+ 06" indicates that the decimal has been moved six places from the right. For the changing of the labels on the y axis to include commas, we used the following function: "scale_y_continuous(labels = scales::comma)", from the "ggplot2" package.

```
Antal_kvinder %>%

ggplot(aes(x = Year, y = Women))+
```

```
geom_point(color = "blue")+
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+
labs(title = "Development in number of women from 1973-2022")+
scale y continuous(labels = scales::comma)
```

It worked and the numbers are now how we would like them to be, clearly in millions



However, we also want to change the color of the dots to make it a part of the same color scale as before. This time we chose another color scale named "Moonrise2" and selected the third color in it. Afterwards we adjusted the size of the points to two.

```
Antal_kvinder %>%

ggplot(aes(x = Year, y = Women))+

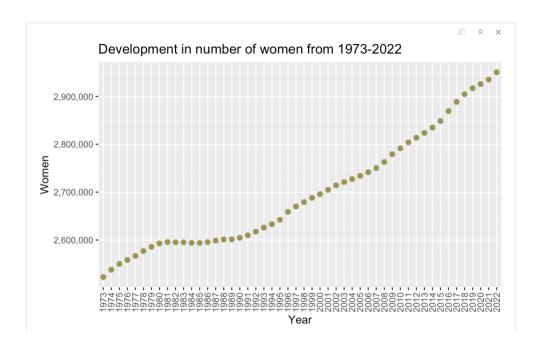
geom_point(color = wes_palette("Moonrise3")[3], size = 2)+

theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+

labs(title = "Development in number of women from 1973-2022")+

scale y continuous(labels = scales::comma)
```

Below you can see the final result of our visualization:



Making a stacked plot visualizing in which weeks the abortions were performed between 1973 and 2022 (everything coded together, Mathilde wrote this part here)

We also wanted to visualize by using a part of the same data from Sundhedsdatastyrelsen²⁴, table 3, which shows in which week of the pregnancy the abortions were performed from 1973 to 2022 and the number of them. The weeks are divided into ranges: 0-7 full weeks, 8-11 full weeks, 12-17 full weeks, 18-22 weeks, 22 full weeks and above as well as Unknown. We wanted to do this to see if there is a trend or a preponderance for one or more intervals and to create an overview to make the data clearer and more manageable and visualize it in a new way. We also hoped to contribute to the new debate about the change of the abortion limit with a graph, so that people have a data visualization that could help them form their opinions on the new abortion limit recommended by Etisk Råd.

As well as with the other data we quickly concluded that it was not tidy for R to work with, so the first task was to alter the data to make that possible.

²⁴ Sundhedsdatastyrelsen and Esundhed (2023) 'Register over Legalt Provokerede Aborter (ABR) <u>https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Provokerede-aborter</u> → look in the excel file, table nr.3

We started by creating a folder for the dataset as well as installing and loading the packages we knew we would need: tidyverse, readxl and janitor. The name of the dataset here is Abortdata_kopi.

Hereafter we loaded the data into R by using the same method as before, pushing the 'Import Dataset' button the top right corner:

```
Abortdata_kopi <- read_excel("data/Abortdata-kopi.xlsx", sheet = "Tabel 3 - Svangerskabslængde", col_names = FALSE, skip = 1)
```

We selected which of the tables in the data we wanted to use, here it was window 3 named 'Svangerskabslængde' and the rows we wanted to skip.

Then we started our needed data transformation, where we selected the rows, we wanted to include in our visualization by cutting the unwanted rows using the function 'slice' and assigned the changes into our dataset by using →:

```
Abortdata kopi %>% slice(2:8) -> Abortdata kopi
```

Here we also wanted to move the years row and make them column names instead of just a row. So, we used the row to names function from janitor again:

```
Abortdata kopi %>% row to names(row number = 1) -> Abortdata kopi
```

Hereafter we needed to make some more changes to the data.

We started by mutating the data to make it the same data type by using the mutate_if function. We wanted to convert it from numeric into character because we wanted to keep the names of the different stages like '0-7 fulde uger', as it would have deleted these had we done it the other way around by converting the data type from character to numeric.

```
Abortdata kopi %>%
```

```
mutate if(is.numeric, as.character) %>%
pivot longer(cols = matches("\d{4}"),
       names to = "Year",
       values to = "Num of abortion") %>%
rename("Stage" = "År") %>%
mutate(Num of abortion = as.numeric(Num of abortion))-> Abort data tidy
```

We also needed to convert the data from a wide format to a long format, which works best for R in our case for us to be able to create the visualizations²⁵. All the information is packed together in the original dataset, and we want a long format where, for example, '0-7 fulde uger' is present for each year from 1973 to 2022. The result of this function as well as the others are illustrated in the screenshot that follows this section. It is just a small cut out to illustrate the changes.

The pivot longer helps us change the format. To capture all the years and make a column that represents all the years we use a regular expression: ("\\d{4}\") (sæt et link ind om regular expressions²⁶). The d represents digits and the curly brackets and the number 4 captures numbers of four digits, here the years. We turn the data and make two columns with the names to and values to. Cols specifies which columns need to be turned. Names to names the variable stored in the column names, in our case it is the years that are column names. The values to name the variable in the cell values, here it is the number of abortions²⁷. Finally, we have the last column, in the dataset it is named 'År', which originates from the original data, but it actually represents the last values, the weeks. We rename this with the rename() function to "Stage". Lastly, we mutate the Num of abortions into numeric and assign everything into a new version of the dataset called Abort data tidy.

²⁵ Wickham, Hadley et al. (2023) "R for Data Science (2e)" Chapter 5 Data Tidying. https://r4ds.hadley.nz/ ²⁶Library Carpentry (18 october 2023) "Regular expressions" https://librarycarpentry.org/lc-data-intro/01regular-expressions.html

Wickham, Hadley et al. (2023) "R for Data Science (2e)" Chapter 5 Data Tidying. https://r4ds.hadley.nz/

^	Stage	Year ‡	Num_of_abortion
1	0-7 fulde uger	1973	990
2	0-7 fulde uger	1974	4310
3	0-7 fulde uger	1975	4360
4	0-7 fulde uger	1976	4025
5	0-7 fulde uger	1977	3765
6	0-7 fulde uger	1978	3330
7	0-7 fulde uger	1979	2585
8	0-7 fulde uger	1980	2675

After this, we originally tried to make a stacked barplot with ggplot, which worked, but a problem was that the order of the Stage levels was messed up. Therefore, we found the solution, that we needed to convert Stage into a factor, so we could change the order²⁸

We use the count() function to count how many observations we are dealing with per level in Stage.

```
Stage_levels <- c("0-7 fulde uger", "8-11 fulde uger", "12-17 fulde uger", "18-21 fulde uger", "22 fulde uger og derover", "Ukendt")
```

Here we create the new order of the levels with the exact names from the dataset.²⁹

```
Abort_data_tidy %>%

mutate(Stage = factor(Stage, levels = Stage_levels)) -> Abort_data_tidy
```

We need to mutate to indicate that we now want this factor and its order as the order of Stage, to be able to use this new order in our upcoming graph.

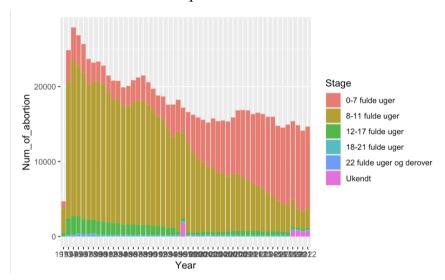
²⁸ Wickham, Hadley et al. (2023) "R for Data Science (2e)" Chapter 16 Factor. https://r4ds.hadley.nz/

²⁹ Wickham, Hadley et al. (2023) "R for Data Science (2e)" Chapter 16 Factor. https://r4ds.hadley.nz/

It is now possible to try and make a ggplot with the wanted order which we do with the following code. We indicate the values we want on the x-and y-axis, here Year and Num_of_abortion. As well as indicating that we want a stacked barplot, and with the stat=identity that the y-values should be used as the height of the bars from their actual size in the dataset³⁰.

```
ggplot(Abort_data_tidy, aes(fill=Stage, y=Num_of_abortion, x=Year)) +
geom_bar(position="stack", stat="identity")
```

We now have a stacked bar plot with the correct order.

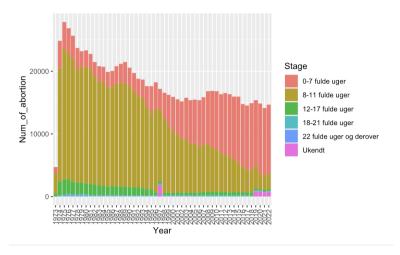


But there are still minor issues to correct such as the years on the x-axis which are messed together and unreadable, so they should be turned, as well as the colors as we want them to be from the same color package as our other graphs.

We started by turning the years on the x-axis with the theme() and 90 means the angle.

³⁰ Zach and Statology. (12 december 2012) "When to use stat=identity in ggplot2 Plots". https://www.statology.org/ggplot-stat-identity/

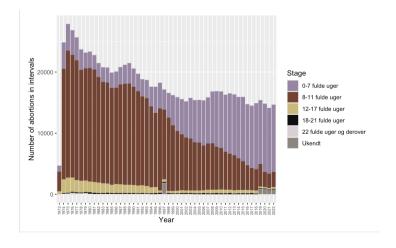
This results in the years being turned 90 degrees, but they are still too close given that we have a lot of values on the x-axis.



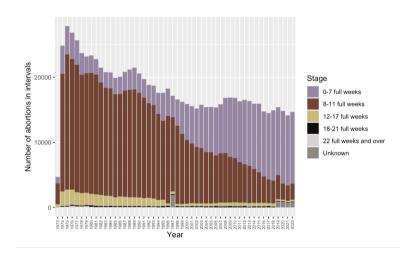
Now it is the task to correct this by adjusting the size of the font and at the same time correct the colors:

```
ggplot(Abort_data_tidy, aes(fill=Stage, y=Num_of_abortion, x=Year)) +
  geom_bar(position="stack", stat="identity")+
  scale_fill_manual(values = wes_palette("IsleofDogs1"))+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1, size = "6"))
  ylab("Number of abortions in intervals")
```

We chose the color scale 'IsleofDogs' from the Wes Anderson color package, adjusted the size of the Year-font by using size = and renamed the y-axis with ylab(). The result was as follows:



Lastly, we wanted to translate all the interval names in the Stage column into english to make it consistent for the whole thing. We did this by including breaks = c() with danish names and labels = c() with the new english names, which gave us the following final result:



Empirical Results (Trine and Mathilde)

The results of our coding process were three visualizations made from data provided by Sundhedsdatastyrelsen and Danmarks Statistik.

First of all, our goal was to illustrate how many abortions that have been performed each year since 1973 to 2022, to see what the development has looked like for the last 50 years since the legalization. It is possible to see that very few were actually performed in the legalization year 1973, but that the number increased sharply in the four following years,

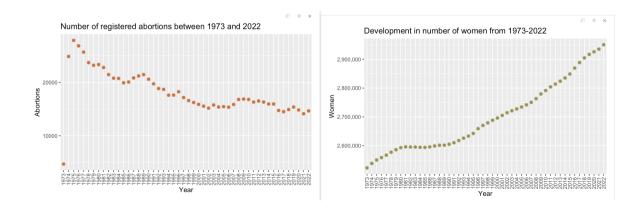
topping in 1975. Since then, the number has decreased steadily, topping a few times in the end of 1980's-beginning of 1990's and around the 2010's. However, given that we are working with big numbers, the fact that the number has increased and decreased in some years doesn't represent drastic changes, but still changes though.

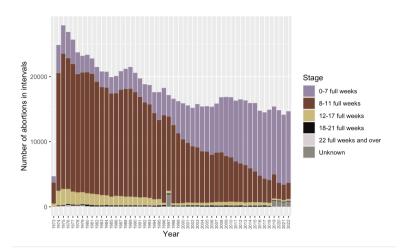
Why the rates increase, and decrease is hard to tell from this graph, and is probably a result of different factors or a coincidence. As well as the reason for abortion, when, thus which week it was performed, or the method.

The most important trend, when we look at the bigger picture, is that the rate has generally <u>dropped</u>. This is probably due to a combination of several factors like more knowledge of relevant parameters, society, political and attitude changes, new methods, and implementations to discover pregnancy, to terminate a pregnancy earlier as well as protection.

The second graph we made on the basis of statistics from Danmarks Statistik was made to compare with the abortion rate and illustrates the development in the number of women from the same time period. As it can be clearly read from the illustration the number of women has increased rather steadily during the last 50 years with almost 500,000. This raise is probably due to a combination of higher life expectancy and a steady birth rate.

We wanted to make this graph, to compare it with the abortion rate, to see if the number of women in Denmark has a relation to the number of abortions. But as we can see, one is increasing and the other is decreasing, so the drop in the abortion rate isn't related to the number of women. The rate hasn't dropped because the number of women has dropped, instead the number of women has increased.





Lastly, we have the last graph, now a stacked bar plot, that illustrates in which weeks the abortions have been performed and is ranged in six intervals, from 0-22 weeks and above and a category for unknown cases. The last three intervals: '18-21 full weeks', '22 full weeks and over' and 'Unknown' represent a small number of cases in graph, whereas the other three '0-7 full weeks', '8-11 full weeks' and '12-17 full weeks' take up the biggest part of the stacked bars in the graph.

The golden color that represents '12-17 weeks', and therefore mostly weeks that surpasses the 12 weeks limit, was more present in the years following the legalization up until the 1990's but has also since been present in the other years, most likely because of the permission to abortion above 12 weeks after consultation.

The two stages that take up the most space and where there has been the biggest change are '0-7 full weeks' and '8-11 full weeks. It stands clearly out that from the beginning most abortions were performed between 8 and 11 weeks in the pregnancy, but since the 2000's and 2010's the tendency has turned around, so that the majority of abortions are now performed earlier in the detected pregnancies, between 0-7 weeks, where it could be assumed that it is done closer to 7 weeks than 0 in most cases, because the pregnancy needs to be possible to test, which can be done earlier today than before, and the time from testing to performing the abortion. But it is still the interval where most abortions are performed today.

But talking about this aspect, it is also important to note, that in general we cannot know from our visualization precisely when in the intervals the abortions have been performed, for example whether it is closer to 8 or 11 weeks. Depending on when it is registered, at the first doctor's appointment or the actual abortion, about which there are differences in our dataset, that will be discussed later on.

It is hard to give an exact reason as to why the abortions are now performed earlier than before, in the 0-7 interval rather than 8-11 weeks, but it is probably due to a complex mixture of several factors like changes in society and healthcare system and the new methods mentioned earlier.

It is also hard to conclude whether there is a cohesion between our graph and the new recommendations from Etisk Råd about extending the limit to 18 weeks and as to whether they *should be* compared. There are a lot of factors that are not readable from the data we are working with, and since it is about people, a kind of soft data, it can be hard to make unambiguous conclusions, because every case is a unique case about a unique situation and person.

Etisk Råd emphasizes some of their reasons, for this ethical topic, about extending the limit that, first of all, there is the consideration for every individual woman's situation, her choices and considerations, while also containing the consideration for the fetus, which only gets bigger the older it gets. Therefore, they emphasize the woman's right to self-determination and her free choice, and that today the permission to get an abortion after 12 weeks is almost given in all cases. Also, that the application a woman has to send in to get an abortion after 12 weeks is assessed by an abort commission who makes the decision, but that they have been criticized for being too opaque and having too big regional differences, which underlines their 18 week limit^{31 32}

A factor that is not a part of our data either is the around 40 women who each year travel to other countries to have their abortion performed after the 12 weeks limit. A discussion that has also become relevant again. It is not the biggest number, but here comes another ethical question, how many women in this category is enough to change the limit³³ ³⁴.

As said before we have not wished to take a stand in this discussion about what we see as the best solution, but wanted to visualize a dataset about abortion and analyze what is readable and unreadable in the graphs and we present some of the arguments of Etisk Råd in

³² Folketinget.dk (2023) "Det Etiske Råds udtalelse om en eventuel ændring af abortgrænsen" https://www.ft.dk/samling/20061/almdel/uer/bilag/12/359392.pdf

³¹ Nielsen, Peter Ingemann (26.september 2023) "Flertal i Etisk Råd anbefaler at hæve abortgrænsen" https://www.altinget.dk/artikel/flertal-i-etisk-raad-anbefaler-at-haeve-abortgrænsen

³³ Heissel, Anders (25.september, 2023) "Svensk jordemoder hjælper danske kvinder med sen abort : Jeg er glad for at kunne hjælpe" https://www.dr.dk/nyheder/indland/svensk-jordemoder-hjælper-danske-kvinder-med-sen-abort-jeg-er-glad-kunne-hjælpe

³⁴ Birk, Christian (16.april 2018) "Danske kvinder tager til udlandet for abort" https://www.kristeligt-dagblad.dk/danmark/danske-kvinder-tager-til-udlandet-abort

favor of raising the limit, so that one can use both and other information as part of one's own decision-making process.

Lessons learned in general while working with this project (Trine and Mathilde)

We decided early in the process that we wanted to work with abortion because it is both a historically as well as a highly relevant topic. We decided to use an already accessible dataset, from both Sundhedsdatastyrelsen and Danmarks Statistik, because we wanted to spend more time in R making the visualizations. Focusing on the most relevant data from the dataset for our case, we got to work on a few but relevant graphs. We chose to translate the text in the datasets to english given this paper is written in english as well.

Given that we also decided to work together, it gave us the possibility to share both frustrations, joys and debate continuously - we coded everything together, and we highly recommend it.

Also given that this is a topic that triggers emotions and opinions and that we chose to partly work with an angle, the recommendations of Etisk Råd, where the changes haven't been implemented, we decided not to take a stand, but make our small contribution to the debate instead.

We have had to find several new packages and functions to be able to tidy the dataset and make the visualizations. Also, how to use other and newly discovered functions in something we had already worked with before like ggplot. All these we found by searching on google, from others' help and recommendations like the R for Data Science book³⁵ and the color package³⁶. We could probably even have spent more time finding other special functions for further details but decided to work consistently with the features we found, to get a better grip on them and use them in different situations and alter them for specific datasets.

It has also been a task to learn how to search for solutions to problems when encountering an error in R. This takes time and sometimes demands several attempts with different very specific wordings to find the correct one, and after that, how to use a provided code in the specific cases.

³⁵ Wickham, Hadley et al. (2023) "R for Data Science (2e)". https://r4ds.hadley.nz/

³⁶ Ram, Karthik (no exact date) "Wes Anderson Palettes" https://github.com/karthik/wesanderson

Critical evaluation (Trine and Mathilde)

Our main data is collected from Sundhedsdatastyrelsen and conclusions based on that should be made with caution.

First of all, provoked abortions done in specialty medical practices are sometimes reported delayed. Therefore, the published numbers from the rencents years might be underestimated in the dataset. Besides there was a transition in the National Patient Registry, LPR2, to LPR3 in 2019, this means that the new abortion registers in 2019 includes abortions from LPR2 up to and including February/March, and then including abortions reported to LPR3, while reports to LPR2 before February/March are included in a previous register. This might result in a period of overestimation of abortions in 2019. During the years there have also been methodological differences in the delimitation of the population, which is another reason why comparison of data from before 2019 should be made with caution.

The definition of the year of a provoked abortion also varies across previously abortion registers and the new registry with data from 2019 and beyond, which may also contribute to an overestimation in 2019. In data from 1973-2018 the abortion year is defined as the last contact in the abortion process and from 2019 and beyond it is defined as the first contact in the abortion process. ³⁷

It is also important to note that in the geom point graph concerning the population of women, we have collected all possible age groups. That has been done knowing that not all of these women, and girls, have been able to bear children, but with the thought in mind, that most of them will be at some time in our time span.

Another critical notion should be that we didn't collect our data ourselves which would only have been possible in an ideal world. Since we don't own a time machine or the amount of time, we had to rely on other people's caution when collecting the data, however, all of our data comes from the state of Denmark, which gives it some kind of authentication. We have also tried to make our data more representative by retrieving data on the number of women from Denmark statistics.

³⁷ Sundhedsdatastyrelsen and Esundhed (2023) 'Register over Legalt Provokerede Aborter (ABR) https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Provokerede-aborter → look in the excel file, first document on information about the data 'Dokumentation'

We can also note that we do not actually cover a 50-year time span, but instead only 49 years until 2022, because the abortion dataset only covers these years given we are still in 2023, the 50 year anniversary of the legalization.

The name of our dataset "Abortion_kopi", concerning our abort data could also have been changed, but it didn't. However, that is not all bad, since we were then able to separate the datasets from one another when we used them working with the different visualizations, thus the geom point and the bar plot, even though it is the same dataset.

Also, our bar plot is based on a huge dataset and might be difficult to read, however we have tried to visualize it as simply as possible and found that it shows a good overview on the discussion on whether or not to move the abortion limit.

Besides we are also well aware that we had a tendency on the subject of abortions and that it isn't possible to be completely objective. However, we have tried to neutralize our tendency by not taking an actual stance in the discussion, and only tried to contribute to the debate with our data visualizations and a review of those, trying to make the debate less about assumptions and more about facts.

Conclusions (Trine and Mathilde)

Our visualizations can contribute to the abortion debate, by giving people a manageable entrance to the understanding of the abortion numbers in Denmark.

From a brief overlook of our data, it can be concluded that the numbers of abortions in Denmark has been generally decreasing since 1973. We can also conclude on the basics of our second visualization that the decrease in abortions has no relation to the number of women in Denmark, since that number has increased rather steadily during the last 50 years with almost 500,000. Finally our third visualization concerning the weeks where abortions were performed between 1973 and 2022, shows clearly that most abortions generally have been within the first 0-7 full weeks or the 8-11 full weeks and that from the beginning most abortions were performed within the first 8-11 weeks of the pregnancy, but ever since the 2000's and 2010's the tendency has turned around, and that the majority of abortions are now instead performed earlier in the detected pregnancies, between 0-7 weeks.

We can also conclude how important tidy data is when working in RStudio and creating visualizations. Even though data can seem tidy in the first place, in our case it was tidy for the human eye, it might not be when working in R, because R has different

Studienummer: 202008537

Studienummer: 202008125

expectations and demands to be able to create visualizations. It can be quite a process to tidy data in R, it takes time and patience, but is also a working process that pays off in the long run and lays the foundation to continuously work with the data. The process also demands different things depending on the data, in this paper and coding it was needed to find new functions and packages in R, to alter the specific data.

Acknowledgements (Trine and Mathilde)

We would like to credit Max Odsbjerg Pedersen for his most generous help when dealing with our problem solving in Rmarkdown, and for his positive attitude toward helping students in need. We would also like to thank Adela Sobotkova for showing us the programs used in this assignment.

References

Datasets:

Sundhedsdatastyrelsen. (2023) "Register over Legalt Provokerede Aborter (ABR)". (https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/graviditet-foedsler-og-boern/abortregisteret (found 8. October 2023)

- Data found directly here: Esundhed (2023). "Provokerede aborter".
 (https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Provokerede-aborter)
 (found 4. October 2023)
 - → specifically table 1 and 3 have been used, as well as the first called
 'Dokumentation'

Danmarks Statistik. Statistikbanken. (2023) "BEFOLK2, Befolkningen 1. januar efter køn og alder (1901-2023).

(https://www.statistikbanken.dk/20021?fbclid=IwAR0OiMFvfwD8sPY4oXTJPqPIgAn7uMn 51ULCrzf43wi5vBxYMwlR-F9HiFY) (found. 8. October 2023)

- → Befolkningen 2.januar efter køn og alder. Kvinder (1973–2022)

- → Choose: 'Kvinder', 'Alder i alt' and mark the years '1973-2022'

E-sundhed (2023) "Anvendelsesvilkår". (https://www.esundhed.dk/Anvendelsesvilkaar) (found 5 October 2023)

Danmarks Statistik (2023) "Kildeangivelse og brug af Danmarks Statistiks logo" (https://www.dst.dk/da/Statistik/brug-statistikken/kildeangivelse) (found 5 October 2023)

Other references:

Birk, Christian (16.april 2018) "Danske kvinder tager til udlandet for abort". Kristeligt Dagblad. (https://www.kristeligt-dagblad.dk/danmark/danske-kvinder-tager-til-udlandet-abort) (found 10 October 2023)

Folketinget.dk (2023) "Det Etiske Råds udtalelse om en eventuel ændring af abortgrænsen" (https://www.ft.dk/samling/20061/almdel/uer/bilag/12/359392.pdf) (found 10 October 2023)

Goodreads (2023): "Abortion quotes" (https://www.goodreads.com/quotes/tag/abortion) (found 9 October 2023)

Heissel, Anders (26.september 2023) "Et flertal i Etisk Råd anbefaler at hæve abortgrænsen til 18 uger". DR.

(https://www.dr.dk/nyheder/indland/et-flertal-i-det-etiske-raad-anbefaler-haeve-abortgraensen-til-18-uger) (found 10 October 2023)

Heissel, Anders (25.september, 2023) "Svensk jordemoder hjælper danske kvinder med sen abort: Jeg er glad for at kunne hjælpe". DR. (https://www.dr.dk/nyheder/indland/svensk-jordemoder-hjaelper-danske-kvinder-med-sen-abort-jeg-er-glad-kunne-hjaelpe) (found. 9. October 2023)

Hytting, Emmelie Sakina (27.Februar 2010) "Tidslinje: Abortens udvikling i Danmark". Etik.dk (https://www.etik.dk/etik/tidslinje-abortens-udvikling-i-danmark) (found 9 October 2023)

Library Carpentry (18 October 2023) "Regular expressions". Library Carpentry. https://librarycarpentry.org/lc-data-intro/01-regular-expressions.html (found 5 September 2023)

Nationalt Center for Etik (2023) "Om det Etiske Råd" (https://nationaltcenterforetik.dk/raad-og-komiteer/det-etiske-raad/om-det-etiske-raad/ (found 8 September 2023)

Nielsen, Peter Ingemann (26.September 2023) "Flertal i Etisk Råd anbefaler at hæve abortgrænsen". Altinget. (https://www.altinget.dk/artikel/flertal-i-etisk-raad-anbefaler-at-haeve-abortgrænsen) (found 10 October 2023)

Ram, Karthik (no exact date) "Wes Anderson Palettes". Github. https://github.com/karthik/wesanderson (found 30 October 2023)

Sørensen, Anne (3.Oktober 2017) "Fri abort, 1973-". Danmarkshistorien.dk (https://danmarkshistorien.dk/vis/materiale/fri-abort) (found 9 October 2023)

Wickham, Hadley et al. (2023) "R for Data Science (2e)". O'Reilly. (https://r4ds.hadley.nz/). Chapters: 5.Data Tidying and 16.Factor

Zach and Statology. (12 December 2012) "When to use stat=identity in ggplot2 Plots". https://www.statology.org/ggplot-stat-identity/ (found 24 October 2023)

Table 1 – Software metadata (Trine and Mathilde)

N	Software metadata	Our metadata
r	description	
S	Current software version	RStudio, version 2023.09.0+463
1		
S	Permanent link to Github	
2	repository where you put	
	your script or R project	
S	Legal Software License	Our project is licensed under Creative Commons Legal
3		Code: CC0 1.0 Universal
S	Computing platform /	3 year old Macbook Pro from 2020 version 13.6 macOS
4	Operating System	Ventura.
S	Installation requirements &	Excel (to open the original dataset if needed)
5	dependencies for software	You need to have one of the newest versions of RStudio
	not used in class	installed.
		When working in RStudio you would need following
		packages: tidyverse, readxl, janitor, wesanderson
S	If available Link to software	None needed other than RStudio (RMarkdown) and Excel
6	documentation for special	
	software	
S	Support email for questions	trinelarsen99@icloud.com / 202008125@post.au.dk
6		mathhalb@icloud.com / 202008537@post.au.dk

Table 2 – Data metadata

Nr	Metadata description	
D1	Abortdata.xlsx	Dataset, in excel, from Sundhedsstyrelsen and their online
		platform Esundhed on abortion in Denmark from 1973 and
		forward. It contains four different tables, "Tabel 1-4". We
		have focused on and used only table 1 on The Number of
		Provoked Abortions from 1973-2022 (Antal provokerede
		aborter, 1973-2022) and table 3 on The Number of Abortions

		1 T 4 CD C 1072 2022 (A 4 1 1 1
		by Length of Pregnancy from 1973-2022 (Antal provokerede
		aborter fordelt på svangerskabslængde, 1973-2022)
		We have found the dataset here:
		https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-
		de-nationale-sundhedsregistre/graviditet-foedsler-og-
		<u>boern/abortregisteret</u> . Click on the link under the headline:
		ABR-data på eSundhed.dk →
		https://www.esundhed.dk/Emner/Graviditet-foedsler-og-
		boern/Provokerede-aborter and it is the excel file here.
D2	Antal_kvinder.xlsx	Dataset, in excel, from Danmarks Statistik on the number of
		women in Denmark from 1973-2022. The table can be found
		by going to Statistikbanken on Danmarks Statistik / dst.dk.
		Click on 'Borgere' and then on the link "BEFOLK2,
		Befolkningen 2.januar efter køn og alder (1901-2023).
		https://www.statistikbanken.dk/20021?fbclid=IwAR0OiMFv
		fwD8sPY4oXTJPqPIgAn7uMn5lULCrzf43wi5vBxYMwlR-
		<u>F9HiFY</u> . Here run the document by choosing 'Kvinder',
		'Alder i alt' and select the years '1973-2022'.
D3	Abortdata_kopi	This is just a copy of document 1 in this table, only focusing
		on table 3 instead of table 1 in the excel dataset. Information
		about both is in the first box.