Statistics with R

# Introduction

Studies show that Age and Marital status influences affect the rate of conception and the decision of women to either commit abortion or go to maternity. Most teenager and adult have an active sexual life which spans among various age groups, however, when it comes to decisions about pregnancy and abortion, younger women are significantly more likely than older women to have an abortion, while older women are substantially less likely than their younger counterparts to have children.

In states and New York City, the majority of women who had abortions () were in their s. About three-in-ten () were in their s; teens ages to accounted for ; and women in their s accounted for . The vast majority of women who had abortions were unmarried (), while married women accounted for , according to the CDC’s Abortion Surveillance report.

# Loading packages

The first phase of data science is to import the data in R. In this case, we will import many packages that would enable us to do.

if(!require("install.load")){  
 install.packages("install.load")  
}  
  
install.load::install\_load(c("tidyverse", "janitor", "readxl", "openxlsx", "scales"))  
  
theme\_set(theme\_bw()) # ggplot theme set to theme\_bw()

# Downloading data into local directory

We used download.file() function to download the data from the [ONS](https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/datasets/conceptionstatisticsenglandandwalesreferencetables/2020/conceptions2020workbook.xlsx) website.

link <- "https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/datasets/conceptionstatisticsenglandandwalesreferencetables/2020/conceptions2020workbook.xlsx"  
  
download.file(link, "conceptions2020.xlsx", mode = "wb")

# Reading the data

We then import data from the workbook and select the most important variables that would be needed for analysis from the data.

conception <- read.xlsx("conceptions2020.xlsx", sheet = "1a", startRow = 8, sep.names = " ") %>% # Select variables needed for the analysis  
 select(  
c("Year of conception", "All ages Number of conceptions", "Under 16 Number of conceptions", "Under 18 Number of conceptions", "Under 20 Number of conceptions", "20 to 24 Number of conceptions", "25 to 29 Number of conceptions", "30 to 34 Number of conceptions", "35 to 39 Number of conceptions", "40 and over Number of conceptions"))

We can see the information about the data by using:

conception %>%   
 head()

Year of conception All ages Number of conceptions  
1 2020 817515  
2 2019 821089  
3 2018 839043  
4 2017 847204  
5 2016 863106  
6 2015 876934  
 Under 16 Number of conceptions Under 18 Number of conceptions  
1 2085 12576  
2 2430 14857  
3 2413 15644  
4 2517 16740  
5 2821 18086  
6 3466 20351  
 Under 20 Number of conceptions 20 to 24 Number of conceptions  
1 42093 138373  
2 48399 145410  
3 51089 151520  
4 52966 155504  
5 56157 161270  
6 59815 168391  
 25 to 29 Number of conceptions 30 to 34 Number of conceptions  
1 224959 248528  
2 228893 239353  
3 235430 242210  
4 240352 240799  
5 247222 242592  
6 249965 244881  
 35 to 39 Number of conceptions 40 and over Number of conceptions  
1 132113 31438  
2 128898 30015  
3 129577 29202  
4 128692 28793  
5 127106 28759  
6 124692 29190

conception %>%   
 dim()

[1] 31 10

conception %>%   
 names()

[1] "Year of conception" "All ages Number of conceptions"   
 [3] "Under 16 Number of conceptions" "Under 18 Number of conceptions"   
 [5] "Under 20 Number of conceptions" "20 to 24 Number of conceptions"   
 [7] "25 to 29 Number of conceptions" "30 to 34 Number of conceptions"   
 [9] "35 to 39 Number of conceptions" "40 and over Number of conceptions"

The conceptions data is a yearly data and it runs from to . The trend of conception by year is shown in [Figure 1](#fig-trend-line) for all ages and it can be seen that the rate of conception is gradually decreasing from till . Also, conception for women in under is higher compared to women in under .

conception %>%  
 ggplot(aes(x = `Year of conception`, y = `All ages Number of conceptions`)) +  
 geom\_line() +  
 scale\_y\_continuous(labels = label\_number(suffix = "K", scale = 1e-3)) +  
 labs(y = "Number of conception", x = "Year")

|  |
| --- |
| Figure 1: Number of conceptions by year. |

The next thing now is to explore the rate of conceptions by age group per year. We need to reshape the data in order to achieve this.

conception\_reshape <- conception %>%  
 pivot\_longer(cols = -(1:2), names\_to = "age\_group", values\_to = "conception") %>%  
 mutate(age\_group = str\_remove\_all(age\_group, "Number.\*")) %>%  
 select(-2)

The reshape data looks like this:

conception\_reshape %>%   
 head()

# A tibble: 6 × 3  
 `Year of conception` age\_group conception  
 <dbl> <chr> <dbl>  
1 2020 "Under 16 " 2085  
2 2020 "Under 18 " 12576  
3 2020 "Under 20 " 42093  
4 2020 "20 to 24 " 138373  
5 2020 "25 to 29 " 224959  
6 2020 "30 to 34 " 248528

In [Figure 2](#fig-conception-reshape), the number of conceptions per year grows for women in the to age group compared to other groups.

conception\_reshape %>%   
 ggplot(aes(x = `Year of conception`, y = conception, col = age\_group)) + geom\_line() + labs(y = "Number of conceptions", col = "Age group") + scale\_y\_continuous(breaks = seq(0, 300000, 50000), labels = label\_number(suffix = "K", scale = 1e-3))

|  |
| --- |
| Figure 2: The trend analysis of number of conceptions by age group per year. |

# Teenage pregnancy

In this section, we will examine conceptions leading to maternities and those that are terminated by abortion for the teenagers in the teenage pregnancy data shown below:

teenage\_conceptions <- read.xlsx("conceptions2020.xlsx", sheet = "1b", startRow = 8, sep.names = " ") %>% # Select variables needed for the analysis  
 select(  
 c("Year of conception", "Under 14 Number of conceptions leading to maternities", "Under 14 Number of conceptions terminated by abortion", "Age 14 Number of conceptions leading to maternities", "Age 14 Number of conceptions terminated by abortion", "Age 15 Number of conceptions leading to maternities", "Age 15 Number of conceptions terminated by abortion", "Under 16 Number of conceptions leading to maternities", "Under 16 Number of conceptions terminated by abortion", "Age 16 Number of conceptions leading to maternities", "Age 16 Number of conceptions terminated by abortion", "Age 17 Number of conceptions leading to maternities", "Age 17 Number of conceptions terminated by abortion", "Age 18 Number of conceptions leading to maternities", "Age 18 Number of conceptions terminated by abortion", "Age 19 Number of conceptions leading to maternities", "Age 19 Conceptions terminated by abortion"))

teenage\_conceptions %>%   
 head()

Year of conception Under 14 Number of conceptions leading to maternities  
1 2020 29  
2 2019 37  
3 2018 42  
4 2017 40  
5 2016 44  
6 2015 61  
 Under 14 Number of conceptions terminated by abortion  
1 60  
2 73  
3 92  
4 64  
5 84  
6 95  
 Age 14 Number of conceptions leading to maternities  
1 179  
2 225  
3 193  
4 204  
5 223  
6 303  
 Age 14 Number of conceptions terminated by abortion  
1 307  
2 389  
3 375  
4 365  
5 451  
6 517  
 Age 15 Number of conceptions leading to maternities  
1 589  
2 646  
3 691  
4 746  
5 820  
6 1040  
 Age 15 Number of conceptions terminated by abortion  
1 921  
2 1060  
3 1020  
4 1098  
5 1199  
6 1450  
 Under 16 Number of conceptions leading to maternities  
1 797  
2 908  
3 926  
4 990  
5 1087  
6 1404  
 Under 16 Number of conceptions terminated by abortion  
1 1288  
2 1522  
3 1487  
4 1527  
5 1734  
6 2062  
 Age 16 Number of conceptions leading to maternities  
1 1736  
2 1951  
3 2114  
4 2398  
5 2615  
6 2976  
 Age 16 Number of conceptions terminated by abortion  
1 1832  
2 2247  
3 2339  
4 2578  
5 2753  
6 3163  
 Age 17 Number of conceptions leading to maternities  
1 3403  
2 3913  
3 4339  
4 4701  
5 5081  
6 5635  
 Age 17 Number of conceptions terminated by abortion  
1 3520  
2 4316  
3 4439  
4 4546  
5 4816  
6 5111  
 Age 18 Number of conceptions leading to maternities  
1 5925  
2 7237  
3 7802  
4 8381  
5 9020  
6 9553  
 Age 18 Number of conceptions terminated by abortion  
1 6091  
2 6953  
3 7258  
4 7271  
5 7276  
6 7775  
 Age 19 Number of conceptions leading to maternities  
1 9130  
2 10196  
3 11171  
4 11622  
5 12715  
6 13179  
 Age 19 Conceptions terminated by abortion  
1 8371  
2 9156  
3 9214  
4 8952  
5 9060  
6 8957

We can see the variables that are useful in the data by using:

teenage\_conceptions %>%   
 dim()

[1] 23 17

teenage\_conceptions %>%   
 names()

[1] "Year of conception"   
 [2] "Under 14 Number of conceptions leading to maternities"  
 [3] "Under 14 Number of conceptions terminated by abortion"  
 [4] "Age 14 Number of conceptions leading to maternities"   
 [5] "Age 14 Number of conceptions terminated by abortion"   
 [6] "Age 15 Number of conceptions leading to maternities"   
 [7] "Age 15 Number of conceptions terminated by abortion"   
 [8] "Under 16 Number of conceptions leading to maternities"  
 [9] "Under 16 Number of conceptions terminated by abortion"  
[10] "Age 16 Number of conceptions leading to maternities"   
[11] "Age 16 Number of conceptions terminated by abortion"   
[12] "Age 17 Number of conceptions leading to maternities"   
[13] "Age 17 Number of conceptions terminated by abortion"   
[14] "Age 18 Number of conceptions leading to maternities"   
[15] "Age 18 Number of conceptions terminated by abortion"   
[16] "Age 19 Number of conceptions leading to maternities"   
[17] "Age 19 Conceptions terminated by abortion"

The next thing now is to explore the rate of conceptions by age group per year. We need to reshape the data in order to achieve this.

teenage\_conceptions\_reshape <- teenage\_conceptions %>%  
 pivot\_longer(cols = -(1), names\_to = "description", values\_to = "statistics") %>%   
 mutate(age\_group = as\_factor(str\_trim(str\_remove\_all(description, "Number.\*|Conceptions.\*"))), conception = str\_c("leading to ", str\_extract(description, "abortion|maternities")), .after = 1) %>% select(-description)

The reshape data looks like this:

teenage\_conceptions\_reshape %>%   
 head()

# A tibble: 6 × 4  
 `Year of conception` age\_group conception statistics  
 <dbl> <fct> <chr> <dbl>  
1 2020 Under 14 leading to maternities 29  
2 2020 Under 14 leading to abortion 60  
3 2020 Age 14 leading to maternities 179  
4 2020 Age 14 leading to abortion 307  
5 2020 Age 15 leading to maternities 589  
6 2020 Age 15 leading to abortion 921

From [Figure 3](#fig-teenage-pregnancy) below there is a decrease in conception in the last years (-). Conception in teenagers is higher with kids in their late teens compare to those in mid teens. Which explain that as teens approaches 18 years there is higher rate of conception. In addition, more teenagers in mid teens are more likely to abort compare to those in their late teens. Which mean as teenagers approach age 18 they are more willing to go through maternity.

teenage\_conceptions\_reshape %>%   
 ggplot(aes(x= `Year of conception`, y = statistics, fill = conception)) + geom\_col(position = position\_dodge()) + facet\_wrap(~age\_group) + labs(y = "Number of conceptions", x = "Year", fill = "Conception", caption = "Source: ONS") + scale\_y\_continuous( labels = label\_number(suffix = "K", scale = 1e-3))

|  |
| --- |
| Figure 3: Conception by teenagers. |

# Conception leading to maternities and abortion

## by marriage

conception\_by\_marriage <- read.xlsx("conceptions2020.xlsx", sheet = "2", startRow = 8, sep.names = " ") %>% # Select variables needed for the analysis  
 select(  
 c("Year of conception", "Under 18 Number of conceptions", "Under 18 Percentage of conceptions leading to abortion", "Under 20 Number of conceptions", "Under 20 Percentage of conceptions leading to abortion", "Age 20 to 24 Number of conceptions", "Age 20 to 24 Percentage of conceptions leading to abortion", "Age 25 to 29 Number of conceptions", "Age 25 to 29 Percentage of conceptions leading to abortion", "Age 30 to 34 Number of conceptions", "Age 30 to 34 Percentage of conceptions leading to abortion", "Age 35 to 39 Number of conceptions", "Age 35 to 39 Percentage of conceptions leading to abortion", "Age 40 and over Number of conceptions", "Age 40 and over Percentage of conceptions leading to abortion")  
 )

We can see the information about variables that are useful in the data by using:

conception\_by\_marriage %>%   
 dim()

[1] 23 15

conception\_by\_marriage %>%   
 names()

[1] "Year of conception"   
 [2] "Under 18 Number of conceptions"   
 [3] "Under 18 Percentage of conceptions leading to abortion"   
 [4] "Under 20 Number of conceptions"   
 [5] "Under 20 Percentage of conceptions leading to abortion"   
 [6] "Age 20 to 24 Number of conceptions"   
 [7] "Age 20 to 24 Percentage of conceptions leading to abortion"   
 [8] "Age 25 to 29 Number of conceptions"   
 [9] "Age 25 to 29 Percentage of conceptions leading to abortion"   
[10] "Age 30 to 34 Number of conceptions"   
[11] "Age 30 to 34 Percentage of conceptions leading to abortion"   
[12] "Age 35 to 39 Number of conceptions"   
[13] "Age 35 to 39 Percentage of conceptions leading to abortion"   
[14] "Age 40 and over Number of conceptions"   
[15] "Age 40 and over Percentage of conceptions leading to abortion"

The reshape data looks like this:

conception\_by\_marriage\_maternity\_abortion <- function() {  
 conception\_by\_marriage = conception\_by\_marriage %>%   
 clean\_names()  
   
 maternity\_abortion = conception\_by\_marriage %>%   
 mutate(  
 under\_18\_number\_of\_conceptions\_leading\_to\_abortion = under\_18\_number\_of\_conceptions \* under\_18\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 under\_18\_number\_of\_conceptions\_leading\_to\_maternity = under\_18\_number\_of\_conceptions - under\_18\_number\_of\_conceptions\_leading\_to\_abortion,  
  
   
 under\_20\_number\_of\_conceptions\_leading\_to\_abortion = under\_20\_number\_of\_conceptions \* under\_20\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 under\_20\_number\_of\_conceptions\_leading\_to\_maternity = under\_20\_number\_of\_conceptions - under\_20\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion = age\_20\_to\_24\_number\_of\_conceptions \* age\_20\_to\_24\_percentage\_of\_conceptions\_leading\_to\_abortion/100,  
   
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity = age\_20\_to\_24\_number\_of\_conceptions - age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion = age\_25\_to\_29\_number\_of\_conceptions \* age\_25\_to\_29\_percentage\_of\_conceptions\_leading\_to\_abortion/100,  
   
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity = age\_25\_to\_29\_number\_of\_conceptions - age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion = age\_30\_to\_34\_number\_of\_conceptions \* age\_30\_to\_34\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity = age\_30\_to\_34\_number\_of\_conceptions - age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion,   
   
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion = age\_35\_to\_39\_number\_of\_conceptions \* age\_35\_to\_39\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity = age\_35\_to\_39\_number\_of\_conceptions - age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion,   
   
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion = age\_40\_and\_over\_number\_of\_conceptions \* age\_40\_and\_over\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity = age\_40\_and\_over\_number\_of\_conceptions - age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion   
   
 )  
   
 maternity\_abortion %>%   
 mutate(across(.cols = everything(), ~round(.))) %>%   
 select(-c(2:15)) %>% mutate(  
 marital\_status = "within marriage or civil partnership"  
 )  
   
}

conception\_by\_marriage\_maternity\_abortion() %>%   
 head()

year\_of\_conception under\_18\_number\_of\_conceptions\_leading\_to\_abortion  
1 2020 26  
2 2019 36  
3 2018 33  
4 2017 34  
5 2016 29  
6 2015 21  
 under\_18\_number\_of\_conceptions\_leading\_to\_maternity  
1 75  
2 97  
3 133  
4 111  
5 167  
6 166  
 under\_20\_number\_of\_conceptions\_leading\_to\_abortion  
1 158  
2 167  
3 166  
4 150  
5 141  
6 157  
 under\_20\_number\_of\_conceptions\_leading\_to\_maternity  
1 919  
2 1114  
3 1266  
4 1278  
5 1439  
6 1476  
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion  
1 1700  
2 1600  
3 1587  
4 1562  
5 1618  
6 1736  
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity  
1 16196  
2 18405  
3 19863  
4 20441  
5 22182  
6 23790  
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion  
1 5714  
2 6108  
3 6026  
4 6101  
5 6159  
6 6488  
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity  
1 71505  
2 79923  
3 83912  
4 89226  
5 96483  
6 101638  
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion  
1 10790  
2 10826  
3 10386  
4 10130  
5 9870  
6 9485  
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity  
1 120796  
2 122833  
3 128096  
4 130571  
5 135271  
6 138713  
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion  
1 10326  
2 9637  
3 9266  
4 8755  
5 8339  
6 7972  
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity  
1 62910  
2 63929  
3 65457  
4 66722  
5 67467  
6 67231  
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion  
1 4337  
2 3847  
3 3683  
4 3457  
5 3335  
6 3461  
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity  
1 11265  
2 11603  
3 11413  
4 11707  
5 12035  
6 12200  
 marital\_status  
1 within marriage or civil partnership  
2 within marriage or civil partnership  
3 within marriage or civil partnership  
4 within marriage or civil partnership  
5 within marriage or civil partnership  
6 within marriage or civil partnership

The columns in the data then look like this:

conception\_by\_marriage\_maternity\_abortion() %>%   
 names()

[1] "year\_of\_conception"   
 [2] "under\_18\_number\_of\_conceptions\_leading\_to\_abortion"   
 [3] "under\_18\_number\_of\_conceptions\_leading\_to\_maternity"   
 [4] "under\_20\_number\_of\_conceptions\_leading\_to\_abortion"   
 [5] "under\_20\_number\_of\_conceptions\_leading\_to\_maternity"   
 [6] "age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion"   
 [7] "age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity"   
 [8] "age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion"   
 [9] "age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity"   
[10] "age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion"   
[11] "age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity"   
[12] "age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion"   
[13] "age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity"   
[14] "age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion"   
[15] "age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity"  
[16] "marital\_status"

We need to reshape the data into variables that include year of conception, age group, conception decision, statistics, and marital status.

conception\_by\_marriage <- conception\_by\_marriage\_maternity\_abortion() %>% pivot\_longer(cols = -c(1, 16), names\_to = "description", values\_to = "statistics") %>% mutate(age\_group = str\_trim(str\_remove\_all(description, "\_number.\*|age\_")) %>% str\_replace\_all("\_", " ") , conception\_decision = str\_extract(description, "leading\_to\_abortion|leading\_to\_maternity") %>% str\_replace\_all("\_", " "), .after = 1) %>% select(-description) %>% relocate(statistics, .before = marital\_status)

## Conceptions outside marriage or civil partnership

conception\_outside\_marriage <- read.xlsx("conceptions2020.xlsx", sheet = "3", startRow = 9, sep.names = " ") %>% # Select variables needed for the analysis  
 select(  
 c("Year of conception", "Under 16 Number of conceptions", "Age under 16 Percentage of conceptions leading to abortion",  
 "Age under 18 Number of conceptions", "Age under 18 Percentage of conceptions leading to abortion", "Age under 20 Number of conceptions", "Age under 20 Percentage of conceptions leading to abortion", "Age 20 to 24 Number of conceptions", "Age 20 to 24 Percentage of conceptions leading to abortion", "Age 25 to 29 Number of conceptions", "Age 25 to 29 Percentage of conceptions leading to abortion", "Age 30 to 34 Number of conceptions", "Age 30 to 34 Percentage of conceptions leading to abortion", "Age 35 to 39 Number of conceptions", "Age 35 to 39 Percentage of conceptions leading to abortion", "Age 40 and over Number of conceptions", "Age 40 and over Percentage of conceptions leading to abortion")  
 )

We can see the information about the variables that are useful in the data by using:

conception\_outside\_marriage %>%   
 dim()

[1] 23 17

conception\_outside\_marriage %>%   
 names()

[1] "Year of conception"   
 [2] "Under 16 Number of conceptions"   
 [3] "Age under 16 Percentage of conceptions leading to abortion"   
 [4] "Age under 18 Number of conceptions"   
 [5] "Age under 18 Percentage of conceptions leading to abortion"   
 [6] "Age under 20 Number of conceptions"   
 [7] "Age under 20 Percentage of conceptions leading to abortion"   
 [8] "Age 20 to 24 Number of conceptions"   
 [9] "Age 20 to 24 Percentage of conceptions leading to abortion"   
[10] "Age 25 to 29 Number of conceptions"   
[11] "Age 25 to 29 Percentage of conceptions leading to abortion"   
[12] "Age 30 to 34 Number of conceptions"   
[13] "Age 30 to 34 Percentage of conceptions leading to abortion"   
[14] "Age 35 to 39 Number of conceptions"   
[15] "Age 35 to 39 Percentage of conceptions leading to abortion"   
[16] "Age 40 and over Number of conceptions"   
[17] "Age 40 and over Percentage of conceptions leading to abortion"

The reshape data looks like this:

conception\_outsidemarriage\_maternity\_abortion <- function() {  
 conception\_outside\_marriage = conception\_outside\_marriage %>%   
 clean\_names()  
   
 conception\_outside\_marriage = conception\_outside\_marriage %>%   
 mutate(  
 under\_16\_number\_of\_conceptions\_leading\_to\_abortion = under\_16\_number\_of\_conceptions \* age\_under\_16\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 under\_16\_number\_of\_conceptions\_leading\_to\_maternity = under\_16\_number\_of\_conceptions - under\_16\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 under\_18\_number\_of\_conceptions\_leading\_to\_abortion = age\_under\_18\_number\_of\_conceptions \* age\_under\_18\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 under\_18\_number\_of\_conceptions\_leading\_to\_maternity = age\_under\_18\_number\_of\_conceptions - under\_18\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 under\_20\_number\_of\_conceptions\_leading\_to\_abortion = age\_under\_20\_number\_of\_conceptions \* age\_under\_20\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 under\_20\_number\_of\_conceptions\_leading\_to\_maternity = age\_under\_20\_number\_of\_conceptions - under\_20\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion = age\_20\_to\_24\_number\_of\_conceptions \* age\_20\_to\_24\_percentage\_of\_conceptions\_leading\_to\_abortion/100,  
   
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity = age\_20\_to\_24\_number\_of\_conceptions - age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion = age\_25\_to\_29\_number\_of\_conceptions \* age\_25\_to\_29\_percentage\_of\_conceptions\_leading\_to\_abortion/100,  
   
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity = age\_25\_to\_29\_number\_of\_conceptions - age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion,  
   
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion = age\_30\_to\_34\_number\_of\_conceptions \* age\_30\_to\_34\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity = age\_30\_to\_34\_number\_of\_conceptions - age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion,   
   
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion = age\_35\_to\_39\_number\_of\_conceptions \* age\_35\_to\_39\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity = age\_35\_to\_39\_number\_of\_conceptions - age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion,   
   
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion = age\_40\_and\_over\_number\_of\_conceptions \* age\_40\_and\_over\_percentage\_of\_conceptions\_leading\_to\_abortion / 100,  
   
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity = age\_40\_and\_over\_number\_of\_conceptions - age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion   
   
 )  
   
 conception\_outside\_marriage %>%   
 mutate(across(.cols = everything(), ~round(.))) %>%   
 select(-c(2:17)) %>% mutate(  
 marital\_status = "outside marriage or civil partnership"  
 )  
   
}

conception\_outsidemarriage\_maternity\_abortion() %>%   
 head()

year\_of\_conception under\_16\_number\_of\_conceptions\_leading\_to\_abortion  
1 2020 1282  
2 2019 1522  
3 2018 1483  
4 2017 1526  
5 2016 1733  
6 2015 2061  
 under\_16\_number\_of\_conceptions\_leading\_to\_maternity  
1 796  
2 906  
3 925  
4 988  
5 1085  
6 1397  
 under\_18\_number\_of\_conceptions\_leading\_to\_abortion  
1 6612  
2 8054  
3 8234  
4 8613  
5 9267  
6 10324  
 under\_18\_number\_of\_conceptions\_leading\_to\_maternity  
1 5863  
2 6670  
3 7244  
4 7982  
5 8623  
6 9840  
 under\_20\_number\_of\_conceptions\_leading\_to\_abortion  
1 2512  
2 24030  
3 24580  
4 24738  
5 25487  
6 26938  
 under\_20\_number\_of\_conceptions\_leading\_to\_maternity  
1 2404  
2 23088  
3 25077  
4 26800  
5 29090  
6 31244  
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion  
1 50962  
2 52545  
3 51378  
4 50063  
5 49352  
6 49574  
 age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity  
1 69515  
2 72860  
3 78692  
4 83438  
5 88118  
6 93291  
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion  
1 45356  
2 45430  
3 44375  
4 42057  
5 40627  
6 38722  
 age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity  
1 102384  
2 97432  
3 101117  
4 102968  
5 103953  
6 103118  
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion  
1 32627  
2 27656  
3 26056  
4 23335  
5 24558  
6 23881  
 age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity  
1 84315  
2 68038  
3 67672  
4 66763  
5 72893  
6 72802  
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion  
1 17899  
2 16766  
3 15633  
4 14474  
5 13697  
6 12818  
 age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity  
1 40978  
2 38566  
3 39221  
4 38741  
5 37603  
6 36671  
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion  
1 6382  
2 5418  
3 548  
4 4797  
5 4740  
6 4749  
 age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity  
1 9454  
2 9147  
3 948  
4 8832  
5 8649  
6 8780  
 marital\_status  
1 outside marriage or civil partnership  
2 outside marriage or civil partnership  
3 outside marriage or civil partnership  
4 outside marriage or civil partnership  
5 outside marriage or civil partnership  
6 outside marriage or civil partnership

The columns in the data then look like this:

conception\_outsidemarriage\_maternity\_abortion() %>%   
 glimpse()

Rows: 23  
Columns: 18  
$ year\_of\_conception <dbl> 2020, 2019,…  
$ under\_16\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 1282, 1522,…  
$ under\_16\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 796, 906, 9…  
$ under\_18\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 6612, 8054,…  
$ under\_18\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 5863, 6670,…  
$ under\_20\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 2512, 24030…  
$ under\_20\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 2404, 23088…  
$ age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 50962, 5254…  
$ age\_20\_to\_24\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 69515, 7286…  
$ age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 45356, 4543…  
$ age\_25\_to\_29\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 102384, 974…  
$ age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 32627, 2765…  
$ age\_30\_to\_34\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 84315, 6803…  
$ age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 17899, 1676…  
$ age\_35\_to\_39\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 40978, 3856…  
$ age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_abortion <dbl> 6382, 5418,…  
$ age\_40\_and\_over\_number\_of\_conceptions\_leading\_to\_maternity <dbl> 9454, 9147,…  
$ marital\_status <chr> "outside ma…

We need to reshape conception\_outsidemarriage\_maternity\_abortion dataframe into variables that include year of conception, age group, conception decision, statistics, and marital status.

conception\_not\_by\_marriage <- conception\_outsidemarriage\_maternity\_abortion() %>% pivot\_longer(cols = -c(1, 18), names\_to = "description", values\_to = "statistics") %>% mutate(age\_group = str\_trim(str\_remove\_all(description, "\_number.\*|age\_")) %>% str\_replace\_all("\_", " ") , conception\_decision = str\_extract(description, "leading\_to\_abortion|leading\_to\_maternity") %>% str\_replace\_all("\_", " "), .after = 1) %>% select(-description) %>% relocate(statistics, .before = marital\_status)

Now that we have the data for both conceptions leading to maternity and abortion by marriage and out of marriage, we then need to combine the data.

conception\_data <- conception\_by\_marriage %>% rbind(conception\_not\_by\_marriage) %>% mutate(age\_group = factor(age\_group, levels = c("under 16", "under 18", "under 20", "20 to 24", "25 to 29", "30 to 34", "35 to 39", "40 and over")))  
  
conception\_data %>%   
 glimpse()

Rows: 690  
Columns: 5  
$ year\_of\_conception <dbl> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 20…  
$ age\_group <fct> under 18, under 18, under 20, under 20, 20 to 24, …  
$ conception\_decision <chr> "leading to abortion", "leading to maternity", "le…  
$ statistics <dbl> 26, 75, 158, 919, 1700, 16196, 5714, 71505, 10790,…  
$ marital\_status <chr> "within marriage or civil partnership", "within ma…

# Graphical analysis of conception decision by age group and marital status

## Which of the conceptions took place within marriage or civil partnerships and how many conceptions lead to an abortion?

conception\_decision <- conception\_data %>%   
 group\_by(marital\_status, conception\_decision) %>%   
 summarise(total = sum(statistics)) %>% ungroup() %>% filter(marital\_status != "outside marriage or civil partnership")

`summarise()` has grouped output by 'marital\_status'. You can override using  
the `.groups` argument.

conception\_decision

# A tibble: 2 × 3  
 marital\_status conception\_decision total  
 <chr> <chr> <dbl>  
1 within marriage or civil partnership leading to abortion 679220  
2 within marriage or civil partnership leading to maternity 7824936

conception\_data %>%   
 ggplot(aes(x= year\_of\_conception, y = statistics, fill = conception\_decision)) + geom\_col(position = position\_dodge()) + facet\_wrap(~marital\_status) + labs(y = "Number of conceptions", x = "Year", fill = "Conception", caption = "Source: ONS") + scale\_y\_continuous( labels = label\_number(breaks = seq(0, 150000, 50000), suffix = "K", scale = 1e-3)) + theme(legend.position = "bottom")

|  |
| --- |
| Figure 4: Conception decision by marital status. |

The number of conceptions that lead to an abortion within marriage or civil partnerships is 679220

## How did conceptions change over this time period both in and out of wedlock and within certain age categories?

conception\_data %>%   
 filter(age\_group %in% c("under 16", "under 18", "under 20", "20 to 24")) %>%   
 ggplot(aes(x= year\_of\_conception, y = statistics, fill = conception\_decision)) + geom\_col(position = position\_dodge()) + facet\_grid(marital\_status ~ age\_group) + labs(y = "Number of conceptions", x = "Year", fill = "Conception", caption = "Source: ONS") + scale\_y\_continuous( labels = label\_number(breaks = seq(0, 150000, 50000), suffix = "K", scale = 1e-3)) + theme(legend.position = "bottom")

|  |
| --- |
| Figure 5: Conception decision by age and marital status (Age under 16 to 24). |

conception\_data %>%   
 filter(age\_group %in% c( "25 to 29", "30 to 34", "35 to 39", "40 and over")) %>%   
 ggplot(aes(x= year\_of\_conception, y = statistics, fill = conception\_decision)) + geom\_col(position = position\_dodge()) + facet\_grid(marital\_status ~ age\_group) + labs(y = "Number of conceptions", x = "Year", fill = "Conception", caption = "Source: ONS") + scale\_y\_continuous( labels = label\_number(breaks = seq(0, 150000, 50000), suffix = "K", scale = 1e-3)) + theme(legend.position = "bottom")

|  |
| --- |
| Figure 6: Conception decision by age and marital status (Age 25 and above). |