Statistics with R

# Loading packages

In this section, we shall load all the necessary packages to enable us download and import the data.

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 %>%   
 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 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 2010 till 2020. Also, conception for women in under 18 is higher compared to women in under 16.

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 25 to 29 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\_pregnancy <- 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"))

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

teenage\_pregnancy %>%   
 dim()

[1] 23 13

teenage\_pregnancy %>%   
 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"

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\_pregnancy\_reshape <- teenage\_pregnancy %>%  
 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\_pregnancy\_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 20 years (2000-2020). 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\_pregnancy\_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")

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