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.

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 conception by age group per year. We need to reshape our data in order to achieve this.