

## Week 2 practice problems

You will need to load two R packages, `tidyverse` and `kableExtra`, to complete this practice problem. Your task is to reproduce plots and investigate variations and covariations in categorical and continuous variables using the data set `PIDiabetes`, source: [kaggle.com](https://www.kaggle.com).

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the study is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. All patients here are females at least 21 years old of Pima Indian heritage.

The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  2.1.3      v dplyr  0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

1. Formulate questions that you would like to address about the data set. Keep in mind the information in this data set is on female patients at least 21 years old of Pima Indian heritage.
2. Explore the data set graphically and numerically to address your questions. Comment on the variations and covariations in the variables.
3. Present the descriptive statistics for the variables you are most interested in as nice tables.