UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

WST 212 Practical 7 -Data Wrangling

Instructions:

- Complete the questions which follow and save all your code in a single R script, named Practical7.R

Submission 1: Code 💻

- Multiple code submissions are allowed and your autograded results will be available shortly after each submission.
- Ensure all variables are named correctly, as incorrectly named variables will not be awarded any marks. (Remember variable names are case sensitive.)
- Ensure your code does not consist of any syntax errors. If your code produces errors when run, the autograder will not be able to mark it.
- Any code commented out (code is commented out when # is typed in front of it) will be considered rough work and will not be marked.
- Once you have completed your submission, ensure the file is submitted on Gradescope, with the correct file name. The autograder will only be able to grade your submission if you use the correct filename.

Guidelines:

- The datasets required to complete the practical have been provided on ClickUP.
- A template that can be used for this assignment has been provided on ClickUP.
- Remember to assign your code to the variables indicated in this document.

Question 1

Consider the dataset Gapminder data set.

a) Filter the data for the Americas in 2007, deselect all other variables. Save your answer as q1a.
b) Identify the observation with lowest gdp per person. Save your answer as q1b.
c) Identify the observation with highest gdp per person. Save your answer as q1c.
d) Create a new variable called *gdp*, defined as the product of the *population size* and *gdp per person*. Save your answer as q1d.
e) Compute the mean life expectancy (across all observations). Name your new column *mean_life_exp*. Save your answer as q1e.
f) Compute the mean life expectancy for each year. Name your new column *mean_life_exp*. Save your answer as q1f.
g) Identify all observations with above average life expectancy. Count the number of observations per continent. Save your answer as q1g.
h) Create a new column in the dataset, called high_life_expectancy, with a value of 1 if the life expectancy is greater than the mean life expectancy, and a value of 0 otherwise. Save your answer as q1h.

Question 2

Consider the seps data set, showing the general reasons people were admitted to hospital by financial year from July 1993 to June 1998.

a) Restructure the give dataset by pushing dates that are currently displayed as columns into rows. (HINT: You should replace the columns of the years with 2 columns, titled year and value). Save your answer as q2a.

b) Using the same dataset you assigned to q2a, again rearrange the dataset, this time pulling the values in the Fields column into their own columns. (HINT: The columns should be named PatientDays and Separations respectively). Save your answer as q2b.