

# 程序代写代做 CS编程辅导

## MTHM501 Formative Coursework 1



data wrangling

### Question

Load in the file “indicator\_hiv\_estimated\_prevalence% 15-49.csv”. This file contains the estimated HIV **prevalence** in people of aged 15-49 in different countries over time. **Prevalence** is defined here to be the estimated number of people living with HIV per 100 population.

**WeChat: cstutorcs**

1. Produce a **tidy** data set called **gp\_hiv** using the tools in **tidyverse** that we introduced in the week 3 practical. The dataset needs to run from 1991 onwards (there is too much missing data prior to that), and we want to end up with three variables (i.e. columns) Country, year and prevalence. Note that a couple of the years have no values in the data set, and by default R reads these columns in as **character** columns. Hence when you **gather()** the data to create a **prevalence** column, all the numbers will be converted into **characters**. One way to deal with this is to convert the column back into numbers using **as.numeric()**.

**QQ: 749389476**

2. Once you have this tidy dataset, run the following code

```
gp_hiv %>%  
  group_by(Country) %>%  
  summarise(MeanPrevalence = mean(prevalence)) %>%  
  mutate(MeanPrevalence=round(MeanPrevalence,1)) %>%  
  head()
```

and produce a table of the output.

Your report should consist of the code (with explanatory comments- aim for one comment per line of code) used to create ‘gp\_hiv’ and the table of the output. You may use any word processor you choose, but save it to pdf for submission. As a guideline, it is not expected that your code should be longer than one side of A4 (and could be considerably less).