# Session 1 homework task: Answer sheet

```
# Data frame example from our Handout 1. You can easily adapt this for
your new table.

> data_df <- data.frame(name = c('bill', 'jane', 'jacques'), age =
  c(23, 54, 8))


# Creating the data frame ------------------------------------------

# Task: Create a data frame called "Language_Exams" with the information
provided in the table. Note: I have added Student ID column, but this
column is not important for the actual calculations.

> language_exams <- data.frame(student_id = c('Elin', 'Spencer',
  'Crystal', 'Arun', 'Lina', 'Maximilian', 'Leyton', 'Alexandra',
  'Valentina', 'Lola', 'Garfield', 'Lucy', 'Shania', 'Arnold', 'Julie',
  'Michaela', 'Nicholas'), exam_1 = c(93, 89, 75, 52, 34, 50, 46, 62,
  84, 68, 74, 51, 84, 34, 57, 25, 72), exam_2 = c(98, 96, 94, 65, 50, 68,
  58, 77, 95, 86, 89, 70, 90, 50, 67, 37, 90))

> language_exams


# The calculations -------------------------------------------------

# What are the mean scores for exam 1 and exam 2?

> mean(language_exams$exam_1)
> mean(language_exams$exam_2)

# What is the difference between the two means?

> mean(language_exams$exam_2) - mean(language_exams$exam_1)

# What are the mean scores for the two exams if you remove extreme
values (the top and bottom 20%) from each?

> mean(language_exams$exam_1, trim=0.2)
> mean(language_exams$exam_2, trim=0.2)

# Based on the previous step (with outliers removed): What is the
difference between the two means now? Please round the value before
reporting the result.

> round(mean(language_exams$exam_2, trim=0.2) -
  mean(language_exams$exam_1, trim=0.2))
```

```
# Can you do steps 3 and 4 in a single command?

# Long version: Multiple lines: Calculate each mean without outliers,
calculate the difference, round the value
```

```
> mean(language_exams$exam_1, trim=0.2) # Calculate mean scores exam 1
  without outliers
> mean(language_exams$exam_2, trim=0.2) # Calculate mean scores exam 2
  without outliers
> 77.63636 - 62.81818 # Difference btw trimmed mean exam 2 and trimmed
  mean exam 1
> round(14.81818) #Rounds the value
```

```
# Single-command version
```

```
> round(mean(language_exams$exam_2, trim=0.2) -
  mean(language_exams$exam_1, trim=0.2))
```