Question 1

We want to know how grades in MDS are influenced by student and course. Given this statistical question, is the following data tidy or untidy? How would you make the data tidy (you do not need to write any code)? Briefly explain your answer.

```
# A tibble: 5 x 4
  Student DSCI511 DSCI523 DSCI574
  <chr>     <dbl>   <dbl>   <dbl>
1 001          75      89      92
2 002          89      88      89
3 003          72      90      95
4 004          85      81      90
5 005          83      90      88
```

Solution

The data is untidy. There is a variable ("course") split over three columns such that there are multiple observations per row. To make the data tidy, we need to make it longer using `pivot_longer()` to gather the three columns "DSCI511", "DSCI523", "DSCI574" to a single column called "course".

Question 2

The first six rows of the gapminder dataset from the gapminder R package is shown below. Fill out the code below the data to remove the year column from the dataset.

```
  country     continent  year lifeExp      pop gdpPercap
  <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
1 Afghanistan Asia       1952    28.8  8425333      779.
2 Afghanistan Asia       1957    30.3  9240934      821.
3 Afghanistan Asia       1962    32.0 10267083      853.
4 Afghanistan Asia       1967    34.0 11537966      836.
5 Afghanistan Asia       1972    36.1 13079460      740.
```

Solution

gapminder |> select(-c(year))

Question 3

A data file that you want to read into R, called `grades.csv` looks like this:

```
Source: UBC MDS
Description: Block 1 Grades
DSCI511, DSCI521, DSCI551, DSCI542
81,91,100,83
70,89,98,78
76,85,90,86
82,93,95,82
69,81,91,83
Data entered by Tomas Beuzen
```

Fill in the missing code below to successfully read this data into R (assume the `tidyverse` library has already been loaded):

```
grades_data <- read_...("grades.csv", ...)
```

Solution

```
grades_data <- read_csv("grades.csv", skip = 2, n_max = 5)
```

Question 4

Given the dataframes `X` and `Y` below, what kind of join would you perform to produce `Z`? Provide the code to do so.

```
{r
## [1] "X"

## # A tibble: 3 x 2
##   breed           origin
##   <chr>           <chr>
## 1 Golden Retriever United Kingdom
## 2 Poodle          Germany
## 3 Pug             China

## [1] "Y"

## # A tibble: 3 x 4
##   weight color hair   breed
##    <dbl> <chr> <chr>  <chr>
## 1     25 Gold  Medium Golden Retriever
## 2     20 Black Medium Poodle
## 3      8 Brown Short  Pug
```

```
## [1] "Z"

## # A tibble: 3 x 5
##   breed            origin         weight color hair
##   <chr>            <chr>           <dbl> <chr> <chr>
## 1 Golden Retriever United Kingdom     25 Gold  Medium
## 2 Poodle           Germany            20 Black Medium
## 3 Pug              China               8 Brown Short
```

Solution

Either of these answers is acceptable: `inner_join(X, Y)` or `left_join(X, Y)`

Question 5

What is the value of e after running this code:

```
d <- 0
e <- 1

if (e > d & e < 0) {
  e <- 10
} else if (e > d | e < 0) {
  e <- e - 10
} else {
  e + 100
}
```

Solution

-9