

Team members:

In this worksheet we will work with the `planets` data frame from the `hellodatascience` R package, containing data on planets in our solar system:

name	mass	length_of_day	mean_temp	n_moons	ring_system	surface_pressure
Mercury	0.330	4222.6	positive	0	FALSE	0.00
Venus	4.870	2802.0	positive	0	FALSE	92.00
Earth	5.970	24.0	positive	1	FALSE	1.00
Mars	0.642	24.7	negative	2	FALSE	0.01
Jupiter	1898.000	9.9	negative	95	TRUE	NA
Saturn	568.000	10.7	negative	146	TRUE	NA
Uranus	86.800	17.2	negative	28	TRUE	NA
Neptune	102.000	16.1	negative	16	TRUE	NA

- `name`: Name of planet
- `mass`: Mass in 10^{24} kg
- `length_of_day`: Length of day in hours
- `mean_temp`: Whether mean temperature in Celsius is positive or not
- `n_moons`: Number of moons
- `ring_system`: Whether the planet has a set of rings around it
- `surface_pressure`: Surface pressure in bars

Question 1

Describe what the `|>` operator does.

Question 2

Write out the data frame that would be printed from this code.

```
planets |>
  select(name, mass, ring_system)
```

Question 3

Write out the data frame that would be printed from the following code.

```
planets |>
  select(name, mass, ring_system) |>
  filter(ring_system == FALSE)
```

Question 4

Describe how the following code behaves differently from the code in question 3.

```
planets <- planets |>
  select(name, mass, ring_system) |>
  filter(ring_system == FALSE)
```

Question 5

Which planets would be the data frame printed by this code?

```
planets |>
  filter(length_of_day > 24 & surface_pressure <= 1)
```

Question 6

Which planets would be the data frame printed by this code?

```
planets |>
  filter(length_of_day > 24 & surface_pressure <= 1)
```

Question 7

How many planets would be the data frame printed by this code?

```
planets |>
  filter(length_of_day != "negative" | ring_system == FALSE)
```

Question 8

Think about the code from question 7. Why did we have quotations around “negative” but not around FALSE?

Question 9

Identify the issue with this code and correct it.

```
planets |>
  select(!name, !mean_temp)
```

Question 10

Write out the data frame that would be printed from this code.

```
planets |>
  select(name, mass, ring_system) |>
  filter(ring_system == FALSE) |>
  mutate(mass_approx = round(mass))
```

Question 11

What type of variable will `moons_per_mass` be (character, logical, or numeric)?

```
planets |>
  mutate(moons_per_mass = n_moons / mass)
```

Question 12

Identify the value of `mass_category` for each planet resulting from this code.

```
planets |>
  mutate(mass_category = case_when(
    mass < 1 ~ "Small",
    mass < 10 ~ "Medium",
    TRUE ~ "Large"
  ))
```

Question 13

I would like you to write one pipeline that does the following (in order):

1. Subsets the columns to only: `name`, `mass`, `n_moons`, `surface_pressure`.
2. Creates a new variable called `relative_day` which stores: “Short” if the length of a day is less than 24 hours, “Regular” if it exactly equals 24, and “Long” if it is longer than 24 hours.
3. Filters to only planets with “Short” days and more than 50 moons.