

# Warmup: data wrangling across columns

Group members:

**Instructions:** Work with a neighbor on the following activity. I will collect the handout at the end of class, and it will be part of your class participation grade. You will be graded only on effort – it is ok if you don't finish all the questions, or get them all correct.

## starwars data

The `starwars` data (from the `dplyr` package) has information on characteristics (height, mass, birth year, etc.) for many different Star Wars characters. Here are the first few rows and columns of the data:

	name	height	mass	hair_color	skin_color	eye_color	birth_year
1	Luke Skywalker	172	77	blond	fair	blue	19.0
2	C-3PO	167	75	<NA>	gold	yellow	112.0
3	R2-D2	96	32	<NA>	white, blue	red	33.0
4	Darth Vader	202	136	none	white	yellow	41.9
5	Leia Organa	150	49	brown	light	brown	19.0
6	Owen Lars	178	120	brown, grey	light	blue	52.0

1. Without listing any column names explicitly, write code to calculate the median of each numeric variable in the `starwars` dataset.

2. Your friend also writes code for question 1, which returns the following output. Explain why your friend is getting NA for each summary statistic.

```
# A tibble: 1 x 3
  median_height median_mass median_birth_year
    <int>         <dbl>         <dbl>
1         NA         NA         NA
```