dplyr (Part One) - Solutions

Exercises

`mpg`

- Load the `dplyr` package – either by loading the specific package or the `tidyverse` package as a whole.

```
library(tidyverse)
```

- Make sure you have the most recent version of `dpylr` (1.0.2) to ensure smooth coding.

```
sessionInfo()
```

- Load the `ggplot2` package as well if you haven't done so yet. Store the default dataset `mpg` (Fuel economy data from 1999 to 2008 for 38 popular models of cars) in a data frame.

```
#library(ggplot2) #if dplyr loaded specifically
cars <- mpg
```

- Change the name of the following variables: `displ` to displacement, `cyl` to cylinders, `trans` to transmission, `fl` to fueltype.

```
cars <- cars %>%
  rename(displacement = displ,
         cylinders = cyl,
         transmission = trans,
         fueltype = fl)
```

- Drop the column `class`.

```
cars <- cars %>%
  select(-class)
```

- Create a new column `distance_diff`, which shows the difference between cty and hwy. What's the highest difference in absolute terms?

```
cars <- cars %>%
  mutate(distance_diff = cty-hwy)

max(abs(cars$distance_diff))
```

```
## [1] 12
```

- Create another new column `fivegal_distance`, which indicates how far one could drive on the highway with 5 gallons in their tank. How many columns do you have in your dataset by now?

```
cars <- cars %>%
  mutate(fivegal_distance = hwy*5)

ncol(cars)
```

```
## [1] 12
```

- Bonus question: Which car model(s) would get you the furthest on the highway with one gallon in the tank?

```
cars %>%
  filter(hwy == max(hwy)) %>%
  pull(model)
```

```
## [1] "jetta"      "new beetle"
```

starwars

- Load the starwars dataset (part of `dplyr`) and store it in a data frame.

```
starwars <- starwars
```

- Drop all columns after `species`.

```
starwars <- starwars %>%
  select(!films:starships)
```

- Create a new column `height_m`, listing the character's height in meters. Drop `height` afterwards.

```
starwars <- starwars %>%
  mutate(height_m = height/100) %>%
  select(-height)
```

- Keep the following variables: `name`, `mass`, all `_color` variables as well as all numeric variables.

```
starwars <- starwars %>%
  select(name, mass,
         contains("_color"),
         where(is.numeric))
```

- Set the height variable as the first column in your dataset. Put all the `_color`-variables at the end of the dataset. Try to find a neater way to do that than just manually typing all the `_color`-variables. Tip: Check out the `relocate()`-helpfile.

```
starwars <- starwars %>%
  relocate(height_m) %>%
  relocate(contains("_color"), .after = last_col())
```

- Bonus question: Which characters were born more than 100 years before the Battle of Yavin?

```
starwars %>%
  filter(birth_year > 100) %>%
  pull(name)
```

```
## [1] "C-3PO"                 "Chewbacca"             "Jabba Desilijic Tiure"
## [4] "Yoda"                  "Dooku"
```