dpylr (Part Two) - Solutions

Exercises

`mpg`

- Let's start with the `mpg` dataset again. How far would one get with the Chevrolet Corvette (automatic transmission) built in 2008 on the highway with one gallon?

```
## [1] 25
```

- Is there any difference between the different models of the Subaru Forester AWD built in 1999?

```
cars %>%
  filter(model == "forester awd" & year == 1999) %>%
  select(hwy)
```

```
## # A tibble: 2 x 1
##     hwy
##   <int>
## 1    25
## 2    24
```

- Calculate the difference again between the reach in the city and the reach on the highway. How many observations in the dataset do show a difference of more than 5 miles?

```
cars %>%
  mutate(difference = abs(cty-hwy)) %>%
  filter(difference > 5) %>%
  nrow()
```

```
## [1] 157
```

- Let’s look at Toyotas only. More precisely, we are interested in the models built after 2000. How many models could drive more than 20 miles in the city with one gallon only?

```
cars %>%
  filter(manufacturer == "toyota",
         year > 2000,
         cty > 20) %>%
  distinct(model)
```

```
## # A tibble: 3 x 1
##   model
##   <chr>
## 1 camry
## 2 camry solara
## 3 corolla
```

- How many car brands produce models that fall under the category “subcompact”?

```
cars %>%
  filter(class == "subcompact") %>%
  distinct(manufacturer)
```

```
## # A tibble: 5 x 1
##   manufacturer
##   <chr>
## 1 ford
## 2 honda
## 3 hyundai
## 4 subaru
## 5 volkswagen
```

- Which car brand has the lowest average engine displacement (in litres) across all their models built in 2008?

```
cars %>%
  filter(year == 2008) %>%
  group_by(manufacturer) %>%
  summarise(enginedisplacement = mean(displ, na.rm = TRUE)) %>%
  filter(enginedisplacement == min(enginedisplacement))
```

```
## # A tibble: 1 x 2
##   manufacturer enginedisplacement
##   <chr>                     <dbl>
## 1 honda                      1.85
```

- Which SUVs of which car brand get you the farthest on the highway on average?

```
cars %>%
  filter(class == "suv") %>%
  group_by(manufacturer) %>%
  summarise(hwydistance = mean(hwy, na.rm = TRUE)) %>%
  filter(hwydistance == max(hwydistance))
```

```
## # A tibble: 1 x 2
##   manufacturer hwydistance
##   <chr>              <dbl>
## 1 subaru                25
```

starwars

- Back to Starwars. Do female, male or characters with no sex (you can ignore the other categories) have the highest age on average?

```
starwars %>%
  group_by(sex) %>%
  summarise(age = mean(birth_year, na.rm = TRUE)) %>%
  filter(sex != "hermaphroditic" & !is.na(sex))
```

```
## # A tibble: 3 x 2
##   sex      age
##   <chr>  <dbl>
## 1 female  47.2
## 2 male    85.5
## 3 none    53.3
```

- Which eye colour is associated with the lowest average height?

```
starwars %>%
  group_by(eye_color) %>%
  summarise(height = mean(height, na.rm = TRUE)) %>%
  filter(height == min(height, na.rm = TRUE))
```

```
## # A tibble: 1 x 2
##   eye_color height
##   <chr>      <dbl>
## 1 red, blue     96
```

- How many characters weigh either over 100kg or are Droids?

```
starwars %>%
  filter(xor(mass > 100, species == "Droid")) %>%
  nrow()
```

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

- How many homeworlds are listed in the dataset?

```
starwars %>%
distinct(homeworld) %>%
  filter(!is.na(homeworld)) %>%
  nrow()
```

```
## [1] 48
```