

Load packages:

```
library(tidyverse)
```

Copy the data (if mistakes are made, it's easy to reload):

```
mpg2 <- mpg
```

## Exercise 1

Reorder the columns such that `year` is first, `drv` second and the rest as they were:

```
# A tibble: 234 x 11
  year drv manufacturer model displ cyl trans  cty  hwy fl  class
  <int> <chr> <chr>      <chr> <dbl> <int> <chr> <int> <int> <chr> <chr>
1  1999 f      audi      a4     1.8     4 auto~  18   29 p  comp~
2  1999 f      audi      a4     1.8     4 manu~  21   29 p  comp~
3  2008 f      audi      a4     2       4 manu~  20   31 p  comp~
4  2008 f      audi      a4     2       4 auto~  21   30 p  comp~
5  1999 f      audi      a4     2.8     6 auto~  16   26 p  comp~
6  1999 f      audi      a4     2.8     6 manu~  18   26 p  comp~
7  2008 f      audi      a4     3.1     6 auto~  18   27 p  comp~
8  1999 4      audi      a4 q~  1.8     4 manu~  18   26 p  comp~
9  1999 4      audi      a4 q~  1.8     4 auto~  16   25 p  comp~
10 2008 4      audi      a4 q~  2       4 manu~  20   28 p  comp~
# ... with 224 more rows
```

## Exercise 2

### Part 1

Find the mean `hwy` for each `class`:

```
# A tibble: 7 x 2
  class      `mean(hwy)`
  <chr>      <dbl>
1 2seater    24.8
2 compact   28.3
3 midsize   27.3
4 minivan   22.4
5 pickup    16.9
6 subcompact 28.1
7 suv       18.1
```

### Part 2

Find the mean `hwy` for each `class`, but only for 4 wheel drives:

```
# A tibble: 5 x 2
  class      `mean(hwy)`
  <chr>      <dbl>
1 compact   25.8
```

2	midsize	24
3	pickup	16.9
4	subcompact	26
5	suv	18.3

## Exercise 3

### Part 1

Find the number of cars with each transmission type:

```
# A tibble: 10 x 2
  trans    `n()`
  <chr>   <int>
1 auto(av)      5
2 auto(l3)      2
3 auto(l4)     83
4 auto(l5)     39
5 auto(l6)      6
6 auto(s4)      3
7 auto(s5)      3
8 auto(s6)     16
9 manual(m5)   58
10 manual(m6)  19
```

### Part 2

Make a new variable, `trans2`, where `trans` is reduced to only a (`auto(*)`) or m (`manual(*)`):

(Hint: `substr("COWIDUR", 1, 4) == "COWI".`)

Find the number of cars with each `trans2` type:

```
# A tibble: 2 x 2
  trans2 `n()`
  <chr>  <int>
1 a      157
2 m      77
```

### Part 3

Find the mean `hwy` for each `class`, but only for auto transmission:

```
# A tibble: 7 x 2
  class    `mean(hwy)`
  <chr>    <dbl>
1 2seater      24
2 compact    27.7
3 midsize    26.9
4 minivan    22.4
5 pickup     16.7
6 subcompact 28.3
7 suv        17.7
```

## Part 4

Find the mean hwy and mean cty for each class and present it like this:

```
# A tibble: 14 x 3
  class      type value
  <chr>     <chr> <dbl>
1 2seater   hwy    24.8
2 compact  hwy    28.3
3 midsize  hwy    27.3
4 minivan  hwy    22.4
5 pickup   hwy    16.9
6 subcompact hwy    28.1
7 suv      hwy    18.1
8 2seater  cty    15.4
9 compact  cty    20.1
10 midsize cty    18.8
11 minivan cty    15.8
12 pickup  cty     13
13 subcompact cty    20.4
14 suv     cty    13.5
```

## Exercise 4

### Part 1

Find the number of models in each class:

```
# A tibble: 7 x 2
  class      n
  <chr>    <int>
1 2seater     1
2 compact     8
3 midsize     8
4 minivan     1
5 pickup      4
6 subcompact  5
7 suv        13

# A tibble: 7 x 2
  class      n
  <chr>    <int>
1 2seater     1
2 compact     8
3 midsize     8
4 minivan     1
5 pickup      4
6 subcompact  5
7 suv        13
```

### Part 2

Find the number of models in each class (class) and with each transmission type (trans2):

```
# A tibble: 13 x 3
# Groups:   class [7]
  class    trans2    n
  <chr>    <chr> <int>
1 2seater  a       1
2 2seater  m       1
3 compact a       8
4 compact m       8
5 midsize a       8
6 midsize m       5
7 minivan a       1
8 pickup  a       4
9 pickup  m       4
10 subcompact a      5
11 subcompact m      5
12 suv    a      13
13 suv    m       4
```

Check:

```
mpg2 %>% filter(class == "2seater", trans2 == "a")
```

```
# A tibble: 2 x 12
  manufacturer model displ year  cyl trans drv   cty  hwy fl  class
  <chr>         <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 chevrolet   corv~  5.7  1999    8 auto~ r    15   23 p  2sea~
2 chevrolet   corv~  6.2  2008    8 auto~ r    15   25 p  2sea~
# ... with 1 more variable: trans2 <chr>
```

```
mpg2 %>% filter(class == "2seater", trans2 == "m")
```

```
# A tibble: 3 x 12
  manufacturer model displ year  cyl trans drv   cty  hwy fl  class
  <chr>         <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
1 chevrolet   corv~  5.7  1999    8 manu~ r    16   26 p  2sea~
2 chevrolet   corv~  6.2  2008    8 manu~ r    16   26 p  2sea~
3 chevrolet   corv~  7    2008    8 manu~ r    15   24 p  2sea~
# ... with 1 more variable: trans2 <chr>
```

### Part 3

Find the relative frequency of transmission type (`trans2`) for each class (`class`):

```
# A tibble: 13 x 5
# Groups:   class [7]
  class    trans2    n    N    p
  <chr>    <chr> <int> <int> <dbl>
1 2seater  a       2     5 0.4
2 2seater  m       3     5 0.6
3 compact a      24    47 0.511
4 compact m      23    47 0.489
5 midsize a      29    41 0.707
6 midsize m      12    41 0.293
7 minivan a      11    11 1
8 pickup  a      20    33 0.606
9 pickup  m      13    33 0.394
```

```

10 subcompact a      16   35 0.457
11 subcompact m      19   35 0.543
12 suv         a      55   62 0.887
13 suv         m       7   62 0.113

```

Hint: ... %>% group\_by(...) %>% summarise(...) %>% group\_by(...) %>% mutate(...).

## Part 4

Present the result from part 3 above as this:

```

# A tibble: 7 x 3
# Groups:   class [7]
  class      a      m
  <chr>    <dbl> <dbl>
1 2seater  0.4    0.6
2 compact 0.511  0.489
3 midsize 0.707  0.293
4 minivan 1      NA
5 pickup  0.606  0.394
6 subcompact 0.457  0.543
7 suv     0.887  0.113

```

## Exercise 5

Show the correlation between `cty` and `hwy` for each `class` and `year`:

```

# A tibble: 2 x 8
  year `2seater` compact midsize minivan pickup subcompact  suv
  <int>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>   <dbl> <dbl>
1  1999     1     0.955  0.645  0.908  0.840   0.964 0.860
2  2008  0.866  0.894  0.882  1.000  0.984   0.971 0.951

```