

Data wrangling

```
library(tidyverse) # dplyr, tidyr, ggplot2 and friends
```

Data wrangling

Resources:

- <https://dplyr.tidyverse.org/>
- <http://r4ds.had.co.nz/transform.html>

Data

Focuses on when data is given. Can be entered manually:

```
d1 <- tibble(w = mtcars$wt[1:2], h = mtcars$hp[1:2])
d1
```

```
## # A tibble: 2 x 2
##       w     h
##   <dbl> <dbl>
## 1  2.62  110
## 2  2.88  110
```

tibble's (<https://cran.r-project.org/web/packages/tibble/vignettes/tibble.html>) are modern data frames that work well with tidyverse.

“Vanilla”/base R:

```
data.frame(w = mtcars$wt, h = mtcars$hp) # base R
```

```
##       w     h
## 1  2.620 110
## 2  2.875 110
## 3  2.320  93
## 4  3.215 110
## 5  3.440 175
## 6  3.460 105
## 7  3.570 245
## 8  3.190  62
## 9  3.150  95
## 10 3.440 123
## 11 3.440 123
## 12 4.070 180
## 13 3.730 180
## 14 3.780 180
## 15 5.250 205
## 16 5.424 215
## 17 5.345 230
## 18 2.200  66
## 19 1.615  52
## 20 1.835  65
## 21 2.465  97
```

```
## 22 3.520 150
## 23 3.435 150
## 24 3.840 245
## 25 3.845 175
## 26 1.935 66
## 27 2.140 91
## 28 1.513 113
## 29 3.170 264
## 30 2.770 175
## 31 3.570 335
## 32 2.780 109
```

```
tibble(w = mtcars$wt, h = mtcars$hp) # tidyverse
```

```
## # A tibble: 32 x 2
##       w     h
##   <dbl> <dbl>
## 1  2.62  110
## 2  2.88  110
## 3  2.32   93
## 4  3.22  110
## 5  3.44  175
## 6  3.46  105
## 7  3.57  245
## 8  3.19   62
## 9  3.15   95
## 10 3.44  123
## # ... with 22 more rows
```

```
d2 <- tribble(
  ~x, ~y,
  1, 2,
  3, 4
)
d2
```

```
## # A tibble: 2 x 2
##       x     y
##   <dbl> <dbl>
## 1     1     2
## 2     3     4
```

Glue:

```
bind_cols(d1, d2)
```

```
## # A tibble: 2 x 4
##       w     h     x     y
##   <dbl> <dbl> <dbl> <dbl>
## 1  2.62  110     1     2
## 2  2.88  110     3     4
```

```
bind_rows(d1, d1)
```

```
## # A tibble: 4 x 2
##       w     h
##   <dbl> <dbl>
## 1  2.62  110
```

```
## 2 2.88 110
## 3 2.62 110
## 4 2.88 110
```

Dataframes vs tibbles

```
mtcars
as_tibble(mtcars)
```

Data types:

- int stands for integers.
- dbl stands for doubles, or real numbers.
- chr stands for character vectors, or strings.
- fctr stands for factors (categorical variables).
- dtm stands for date-times (a date + a time).
- Others: lgl, date

The pipe operator, %>%

```
filter(mpg, hwy >= 20, hwy <= 26)
```

```
## # A tibble: 87 x 11
##   manufacturer model displ  year  cyl trans drv   cty  hwy fl  cla~
##   <chr>         <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <ch~
## 1 audi         a4      2.8  1999    6 auto~ f    16   26 p  com~
## 2 audi         a4      2.8  1999    6 manu~ f    18   26 p  com~
## 3 audi         a4 q~   1.8  1999    4 manu~ 4    18   26 p  com~
## 4 audi         a4 q~   1.8  1999    4 auto~ 4    16   25 p  com~
## 5 audi         a4 q~   2.8  1999    6 auto~ 4    15   25 p  com~
## 6 audi         a4 q~   2.8  1999    6 manu~ 4    17   25 p  com~
## 7 audi         a4 q~   3.1  2008    6 auto~ 4    17   25 p  com~
## 8 audi         a4 q~   3.1  2008    6 manu~ 4    15   25 p  com~
## 9 audi         a6 q~   2.8  1999    6 auto~ 4    15   24 p  mid~
## 10 audi        a6 q~   3.1  2008    6 auto~ 4    17   25 p  mid~
## # ... with 77 more rows
```

```
select(filter(mpg, hwy >= 20, hwy <= 26), manufacturer, hwy)
```

```
## # A tibble: 87 x 2
##   manufacturer  hwy
##   <chr>        <int>
## 1 audi          26
## 2 audi          26
## 3 audi          26
## 4 audi          25
## 5 audi          25
## 6 audi          25
## 7 audi          25
## 8 audi          25
## 9 audi          24
## 10 audi         25
## # ... with 77 more rows
```

```
# filter(mpg, hwy >= 20, hwy <= 26)
mpg %>% filter(hwy >= 20, hwy <= 26)
```

```
## # A tibble: 87 x 11
##   manufacturer model displ  year  cyl trans drv   cty  hwy fl  cla~
##   <chr>          <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <ch>
## 1 audi          a4      2.8  1999   6 auto~ f    16   26 p  com~
## 2 audi          a4      2.8  1999   6 manu~ f    18   26 p  com~
## 3 audi          a4 q~   1.8  1999   4 manu~ 4    18   26 p  com~
## 4 audi          a4 q~   1.8  1999   4 auto~ 4    16   25 p  com~
## 5 audi          a4 q~   2.8  1999   6 auto~ 4    15   25 p  com~
## 6 audi          a4 q~   2.8  1999   6 manu~ 4    17   25 p  com~
## 7 audi          a4 q~   3.1  2008   6 auto~ 4    17   25 p  com~
## 8 audi          a4 q~   3.1  2008   6 manu~ 4    15   25 p  com~
## 9 audi          a6 q~   2.8  1999   6 auto~ 4    15   24 p  mid~
## 10 audi         a6 q~   3.1  2008   6 auto~ 4    17   25 p  mid~
## # ... with 77 more rows
```

- Each data handling operation performed results in one more level of nested calls.
- The code will quickly become difficult to read
 - Read it from the inner most call and out
 - Operation description (function name) separated from arguments

The pipe operator.

A crucial design decision of the `tidyverse`:

- First argument is a `tibble` (or `data.frame`)
- Subsequent arguments say what to do with the `tibble` (or `data.frame`)
- Always return a `tibble` (or `data.frame`)

This enables the so-called piping operator, `%>%`. Example:

$f(x, \dots) == x \%>\% f(\dots)$

```
sum(sqrt(1:10))
```

```
## [1] 22.46828
```

```
1:10 %>% sqrt() %>% sum()
```

```
## [1] 22.46828
```

```
filter(mpg, hwy >= 20, hwy <= 26) %>%
  select(manufacturer, hwy)
```

```
## # A tibble: 87 x 2
##   manufacturer  hwy
##   <chr>        <int>
## 1 audi          26
## 2 audi          26
## 3 audi          26
## 4 audi          25
## 5 audi          25
## 6 audi          25
## 7 audi          25
```

```
## 8 audi          25
## 9 audi          24
## 10 audi         25
## # ... with 77 more rows

mpg %>%
  filter(hwy >= 20, hwy <= 26) %>%
  select(manufacturer, hwy)

## # A tibble: 87 x 2
##   manufacturer hwy
##   <chr>        <int>
## 1 audi          26
## 2 audi          26
## 3 audi          26
## 4 audi          25
## 5 audi          25
## 6 audi          25
## 7 audi          25
## 8 audi          25
## 9 audi          24
## 10 audi         25
## # ... with 77 more rows
```

Operations

- `filter()`: keep rows matching criteria
- `select()`: pick/reorder columns by name
- `rename()`: rename columns by name
- `arrange()`: reorder rows
- `mutate()`: add new or modify existing variables
- `summarise()`: perform aggregation operations
- `group_by()`: introduce aggregation groupings
- `count()`: a short-hand for `group_by()` and `summarise()`

A ton of helper functions.

Examples

```
mpg %>% select(manufacturer, hwy)
```

```
## # A tibble: 234 x 2
##   manufacturer hwy
##   <chr>        <int>
## 1 audi          29
## 2 audi          29
## 3 audi          31
## 4 audi          30
## 5 audi          26
## 6 audi          26
## 7 audi          27
## 8 audi          26
## 9 audi          25
```

```
## 10 audi                28
## # ... with 224 more rows
```

```
mpg %>% select(class, everything())
```

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

```
mpg %>% select(-manufacturer, -hwy)
```

```
## # A tibble: 234 x 9
##   model      displ  year  cyl trans      drv   cty fl  class
##   <chr>      <dbl> <int> <int> <chr>    <chr> <int> <chr> <chr>
## 1 a4          1.8  1999   4 auto(15) f    18 p  compact
## 2 a4          1.8  1999   4 manual(m5) f    21 p  compact
## 3 a4          2    2008   4 manual(m6) f    20 p  compact
## 4 a4          2    2008   4 auto(av) f    21 p  compact
## 5 a4          2.8  1999   6 auto(15) f    16 p  compact
## 6 a4          2.8  1999   6 manual(m5) f    18 p  compact
## 7 a4          3.1  2008   6 auto(av) f    18 p  compact
## 8 a4 quattro  1.8  1999   4 manual(m5) 4    18 p  compact
## 9 a4 quattro  1.8  1999   4 auto(15) 4    16 p  compact
## 10 a4 quattro  2    2008   4 manual(m6) 4    20 p  compact
## # ... with 224 more rows
```

```
count(mpg, class)
```

```
## # A tibble: 7 x 2
##   class      n
##   <chr>    <int>
## 1 2seater     5
## 2 compact   47
## 3 midsize   41
## 4 minivan   11
## 5 pickup    33
## 6 subcompact 35
## 7 suv       62
```

```
mpg %>% count(class)
```

```
## # A tibble: 7 x 2
##   class      n
##   <chr>    <int>
## 1 2seater     5
## 2 compact   47
```

```
## 3 midsize      41
## 4 minivan     11
## 5 pickup      33
## 6 subcompact  35
## 7 suv         62
```

```
mpg %>% group_by(class) %>% summarise(n = n())
```

```
## # A tibble: 7 x 2
##   class      n
##   <chr>    <int>
## 1 2seater     5
## 2 compact   47
## 3 midsize   41
## 4 minivan   11
## 5 pickup    33
## 6 subcompact 35
## 7 suv      62
```

```
mpg %>%
  group_by(class) %>%
  summarise(
    n = n(),
    displ = mean(displ)) %>%
  arrange(desc(n))
```

```
## # A tibble: 7 x 3
##   class      n displ
##   <chr>    <int> <dbl>
## 1 suv       62  4.46
## 2 compact  47  2.33
## 3 midsize  41  2.92
## 4 subcompact 35  2.66
## 5 pickup   33  4.42
## 6 minivan  11  3.39
## 7 2seater   5  6.16
```

```
mpg %>%
  group_by(manufacturer) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(desc(mean_hwy))
```

```
## # A tibble: 15 x 2
##   manufacturer mean_hwy
##   <chr>         <dbl>
## 1 honda         32.6
## 2 volkswagen    29.2
## 3 hyundai      26.9
## 4 audi          26.4
## 5 pontiac       26.4
## 6 subaru        25.6
## 7 toyota        24.9
## 8 nissan         24.6
## 9 chevrolet     21.9
## 10 ford         19.4
## 11 mercury       18
```

```
## 12 dodge          17.9
## 13 jeep           17.6
## 14 lincoln        17
## 15 land rover    16.5
```

```
mpg %>%
  group_by(cyl, trans) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(cyl, desc(mean_hwy))
```

```
## # A tibble: 27 x 3
## # Groups:   cyl [4]
##   cyl trans      mean_hwy
##   <int> <chr>      <dbl>
## 1     4 auto(l5)        31
## 2     4 auto(s5)        31
## 3     4 auto(av)       30.5
## 4     4 manual(m6)     29.6
## 5     4 manual(m5)     29.3
## 6     4 auto(s6)       28.2
## 7     4 auto(l4)       27.6
## 8     4 auto(l3)       27
## 9     4 auto(s4)       26
## 10    5 auto(s6)       29
## # ... with 17 more rows
```

```
mpg %>%
  group_by(cyl, trans) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(cyl, desc(mean_hwy)) %>%
  identity()
```

```
## # A tibble: 27 x 3
## # Groups:   cyl [4]
##   cyl trans      mean_hwy
##   <int> <chr>      <dbl>
## 1     4 auto(l5)        31
## 2     4 auto(s5)        31
## 3     4 auto(av)       30.5
## 4     4 manual(m6)     29.6
## 5     4 manual(m5)     29.3
## 6     4 auto(s6)       28.2
## 7     4 auto(l4)       27.6
## 8     4 auto(l3)       27
## 9     4 auto(s4)       26
## 10    5 auto(s6)       29
## # ... with 17 more rows
```

```
mpg %>%
  group_by(cyl, trans) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(cyl, desc(mean_hwy)) %>%
  I()
```

```
## # A tibble: 27 x 3
## # Groups:   cyl [4]
```



```
##      cyl trans      mean_hwy
## * <int> <chr>      <dbl>
## 1     4 auto(l5)      31
## 2     4 auto(s5)      31
## 3     4 auto(av)      30.5
## 4     4 manual(m6)    29.6
## 5     4 manual(m5)    29.3
## 6     4 auto(s6)      28.2
## 7     4 auto(l4)      27.6
## 8     4 auto(l3)      27
## 9     4 auto(s4)      26
## 10    5 auto(s6)      29
## # ... with 17 more rows
```

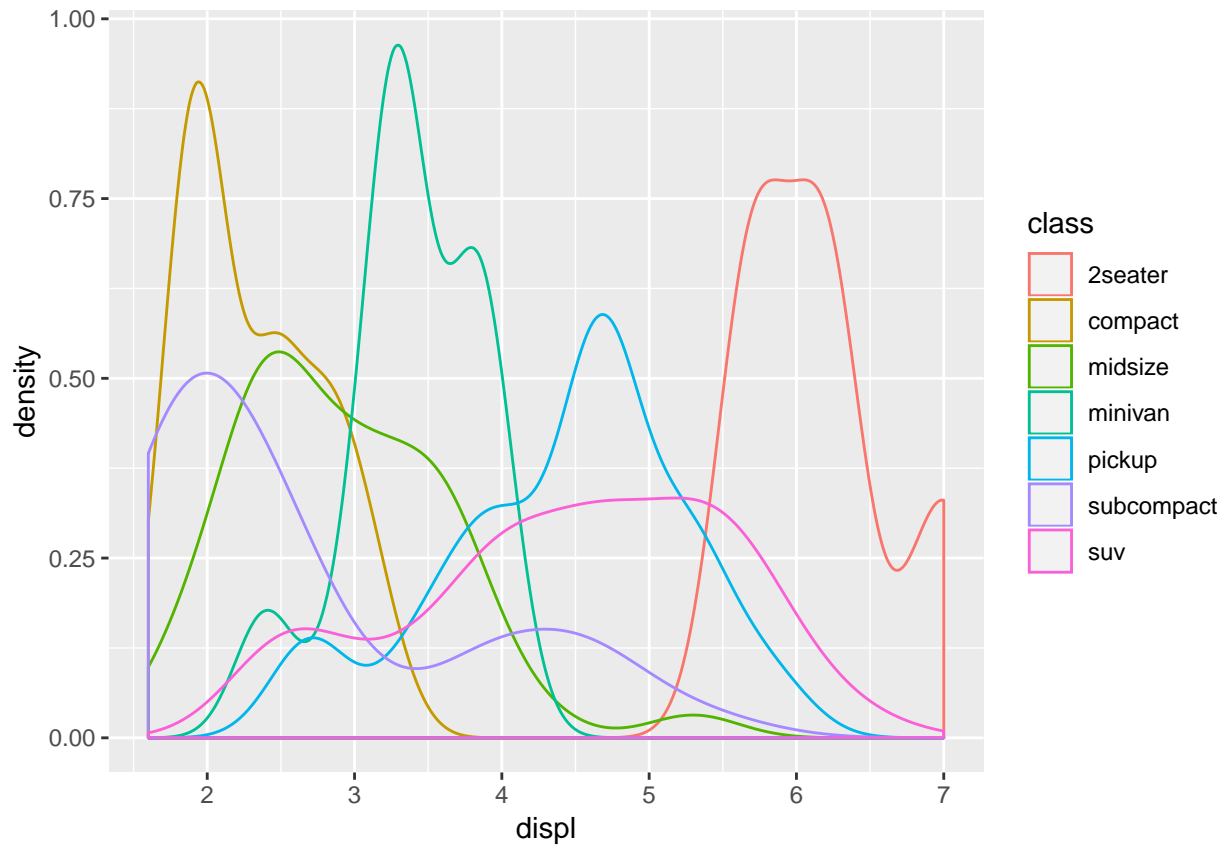
```
mpg %>%
  group_by(cyl, trans) %>%
  summarise(mean_hwy = mean(hwy)) %>%
  arrange(cyl, desc(mean_hwy)) %>%
  I
```

```
## # A tibble: 27 x 3
## # Groups:   cyl [4]
##      cyl trans      mean_hwy
## * <int> <chr>      <dbl>
## 1     4 auto(l5)      31
## 2     4 auto(s5)      31
## 3     4 auto(av)      30.5
## 4     4 manual(m6)    29.6
## 5     4 manual(m5)    29.3
## 6     4 auto(s6)      28.2
## 7     4 auto(l4)      27.6
## 8     4 auto(l3)      27
## 9     4 auto(s4)      26
## 10    5 auto(s6)      29
## # ... with 17 more rows
```

```
mpg %>%
  group_by(class, displ) %>%
  summarise(mean_hwy = mean(hwy))
```

```
## # A tibble: 77 x 3
## # Groups:   class [?]
##   class displ mean_hwy
##   <chr> <dbl> <dbl>
## 1 2seater  5.7   24.5
## 2 2seater  6.2   25.5
## 3 2seater  7     24
## 4 compact  1.8   31
## 5 compact  1.9   44
## 6 compact  2     28.5
## 7 compact  2.2   28
## 8 compact  2.4   29.5
## 9 compact  2.5   27
## 10 compact  2.8   24.7
## # ... with 67 more rows
```

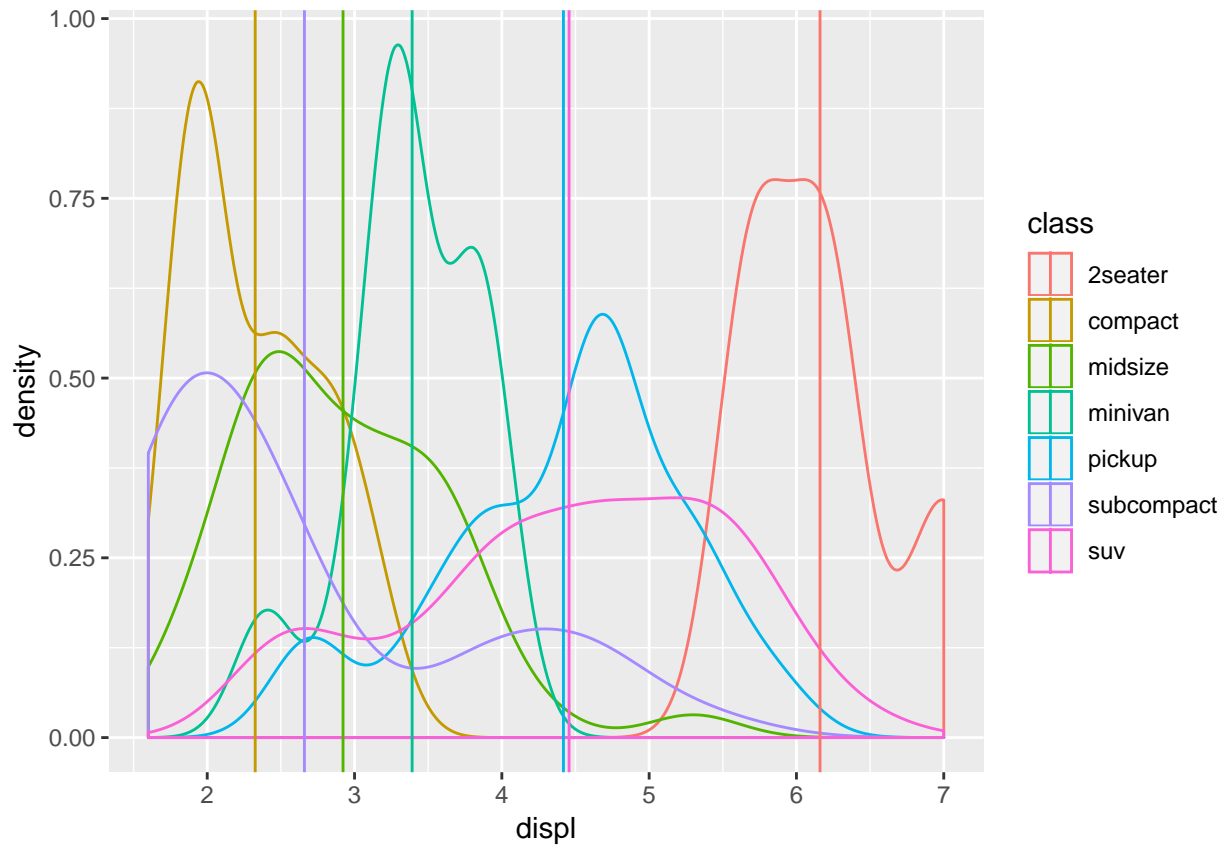
```
ggplot(mpg, aes(displ, color = class)) +
  geom_density()
```



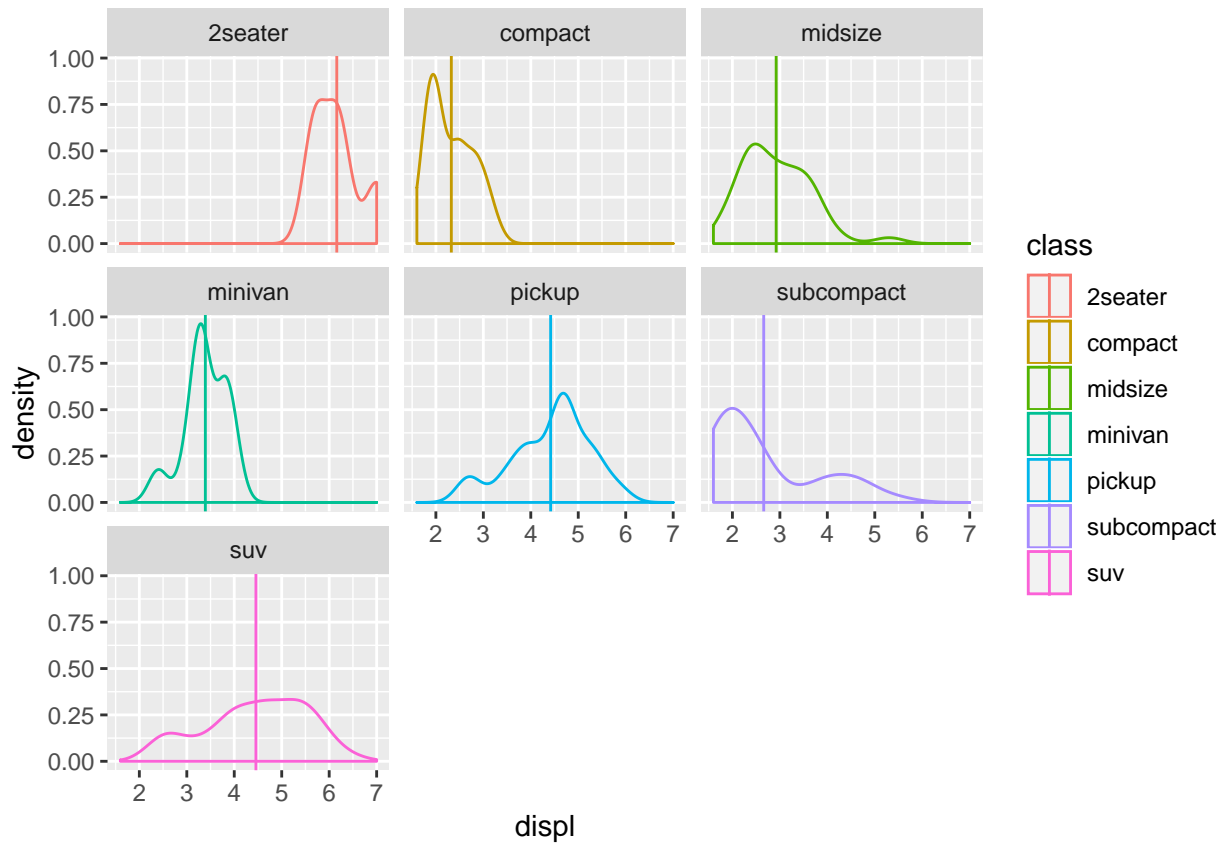
```
d_tmp <- mpg %>%
  group_by(class) %>%
  summarise(displ = mean(displ))
d_tmp
```

```
## # A tibble: 7 x 2
##   class      displ
##   <chr>      <dbl>
## 1 2seater     6.16
## 2 compact     2.33
## 3 midsize     2.92
## 4 minivan     3.39
## 5 pickup     4.42
## 6 subcompact  2.66
## 7 suv         4.46
```

```
ggplot(mpg, aes(displ, color = class)) +
  geom_density() +
  geom_vline(data = d_tmp, aes(xintercept = displ, color = class))
```



```
ggplot(mpg, aes(displ, color = class)) +
  geom_density() +
  geom_vline(data = d_tmp, aes(xintercept = displ, color = class)) +
  facet_wrap(~ class)
```



Long/wide

Wide to long

```
mpg
```

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

```
mpg %>%
```

```
  select(class, hwy, cty)
```

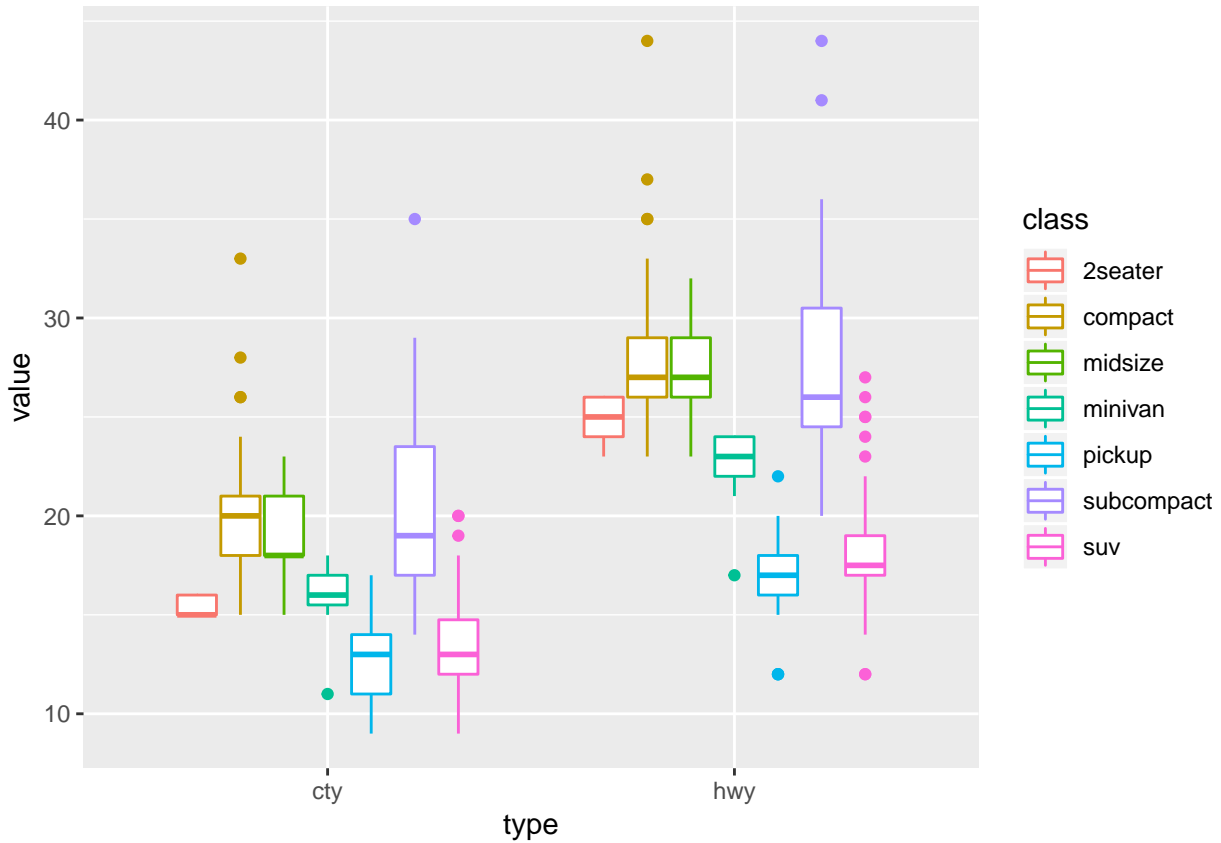
```
## # A tibble: 234 x 3
##   class  hwy  cty
```

```
##   <chr>   <int> <int>
## 1 compact    29    18
## 2 compact    29    21
## 3 compact    31    20
## 4 compact    30    21
## 5 compact    26    16
## 6 compact    26    18
## 7 compact    27    18
## 8 compact    26    18
## 9 compact    25    16
## 10 compact   28    20
## # ... with 224 more rows
```

```
mpg %>%
  select(class, hwy, cty) %>%
  gather(type, value, -class) # gather(type, value, hwy, cty)
```

```
## # A tibble: 468 x 3
##   class  type  value
##   <chr> <chr> <int>
## 1 compact hwy     29
## 2 compact hwy     29
## 3 compact hwy     31
## 4 compact hwy     30
## 5 compact hwy     26
## 6 compact hwy     26
## 7 compact hwy     27
## 8 compact hwy     26
## 9 compact hwy     25
## 10 compact hwy     28
## # ... with 458 more rows
```

```
mpg %>%
  select(class, hwy, cty) %>%
  gather(type, value, -class) %>%
  ggplot(aes(type, value, color = class)) +
  geom_boxplot()
```



Long to wide

```
mpg
```

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

```
mpg %>%
```

```
  mutate(drv = paste0("drv_", drv)) %>%
  select(class, drv, hwy)
```

```
## # A tibble: 234 x 3
##   class  drv   hwy
##   <chr> <chr> <int>
## 1 compact drv_f    29
```

```
## 2 compact drv_f 29
## 3 compact drv_f 31
## 4 compact drv_f 30
## 5 compact drv_f 26
## 6 compact drv_f 26
## 7 compact drv_f 27
## 8 compact drv_4 26
## 9 compact drv_4 25
## 10 compact drv_4 28
## # ... with 224 more rows
```

```
d_tmp <- mpg %>%
  mutate(drv = paste0("drv_", drv)) %>%
  select(class, drv, hwy) %>%
  group_by(drv, class) %>%
  summarise(hwy = mean(hwy))
d_tmp
```

```
## # A tibble: 12 x 3
## # Groups:   drv [?]
##   drv   class    hwy
##   <chr> <chr>    <dbl>
## 1 drv_4 compact  25.8
## 2 drv_4 midsize  24
## 3 drv_4 pickup   16.9
## 4 drv_4 subcompact 26
## 5 drv_4 suv     18.3
## 6 drv_f compact  29.1
## 7 drv_f midsize  27.6
## 8 drv_f minivan  22.4
## 9 drv_f subcompact 30.5
## 10 drv_r 2seater  24.8
## 11 drv_r subcompact 23.2
## 12 drv_r suv     17.5
```

```
d_tmp %>%
  spread(drv, hwy)
```

```
## # A tibble: 7 x 4
##   class      drv_4 drv_f drv_r
##   <chr>    <dbl> <dbl> <dbl>
## 1 2seater    NA    NA  24.8
## 2 compact  25.8  29.1  NA
## 3 midsize   24    27.6  NA
## 4 minivan   NA    22.4  NA
## 5 pickup   16.9  NA    NA
## 6 subcompact 26    30.5  23.2
## 7 suv      18.3  NA    17.5
```

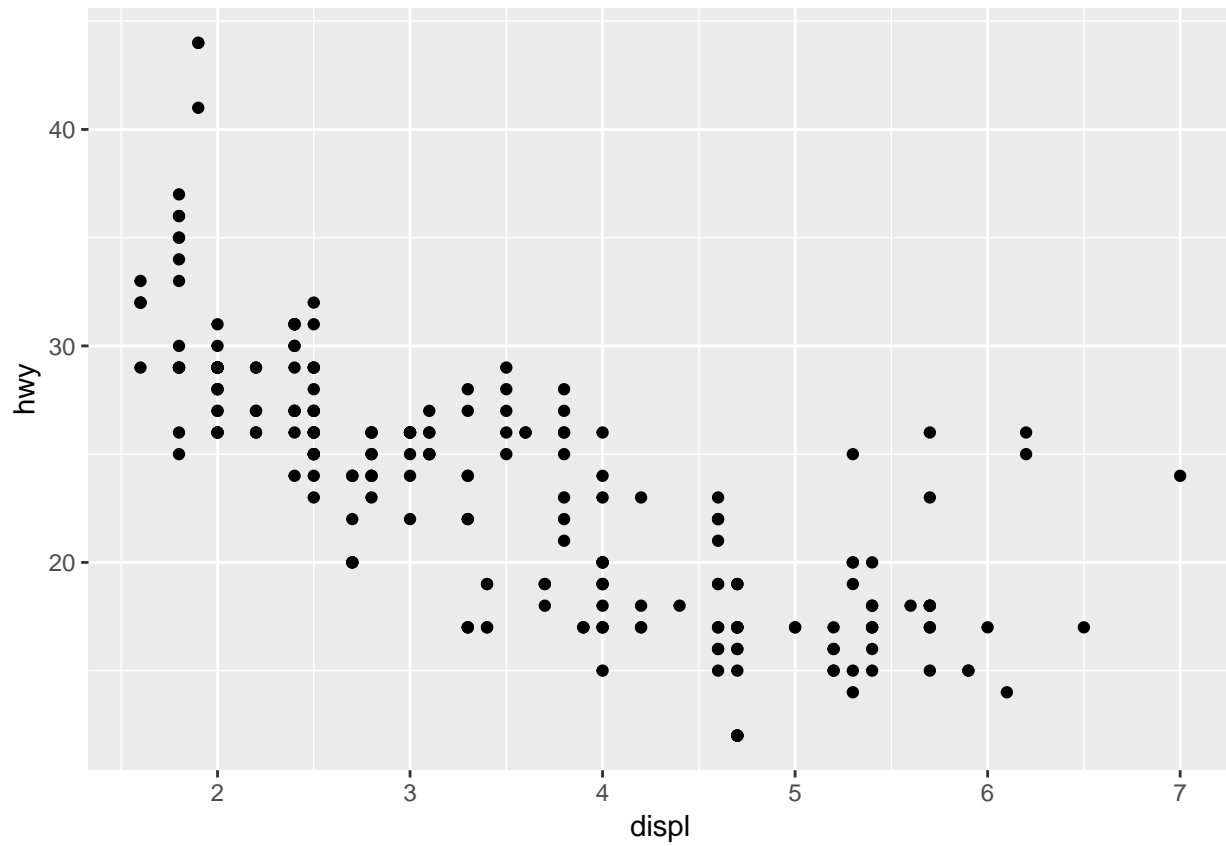
```
d_tmp %>%
  spread(drv, hwy, fill = 0)
```

```
## # A tibble: 7 x 4
##   class      drv_4 drv_f drv_r
##   <chr>    <dbl> <dbl> <dbl>
## 1 2seater    0     0  24.8
```

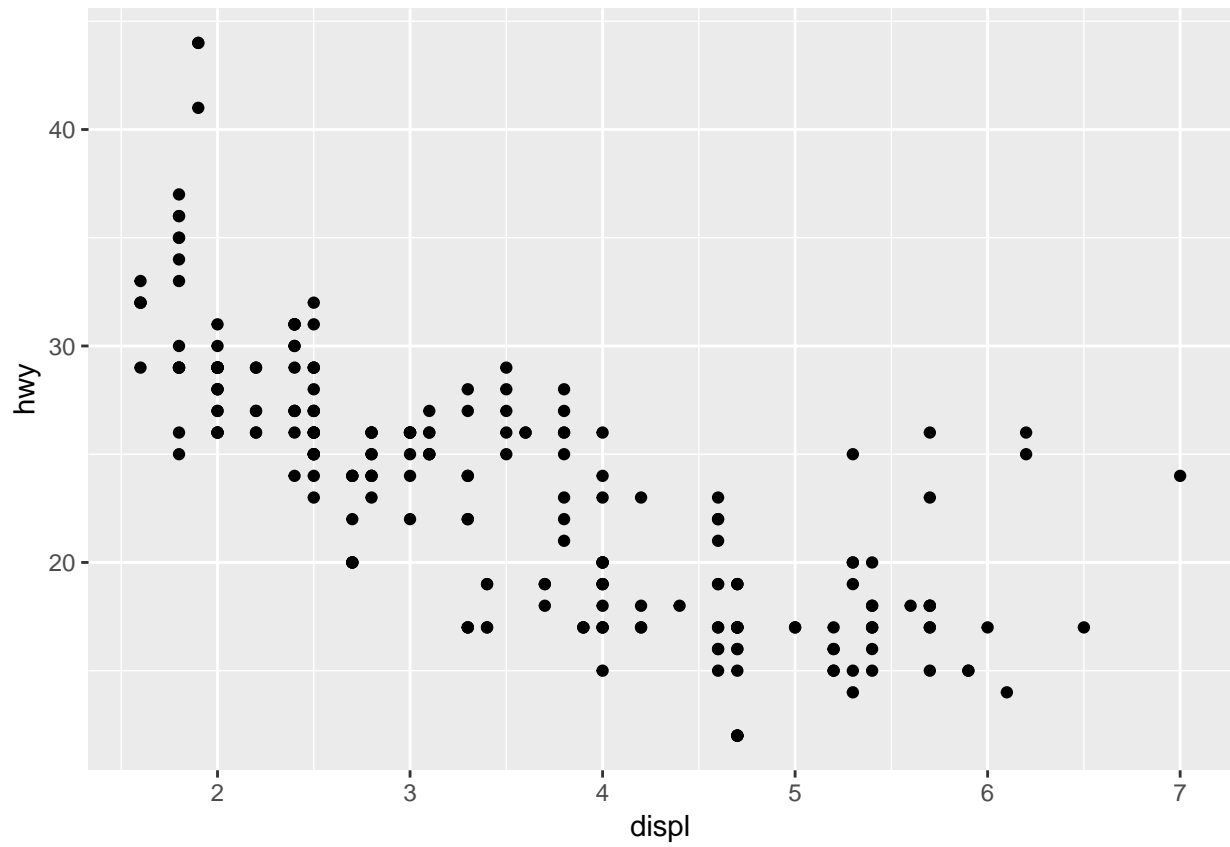
```
## 2 compact      25.8  29.1  0
## 3 midsize      24    27.6  0
## 4 minivan       0    22.4  0
## 5 pickup       16.9   0    0
## 6 subcompact   26    30.5  23.2
## 7 suv          18.3   0    17.5
```

List columns

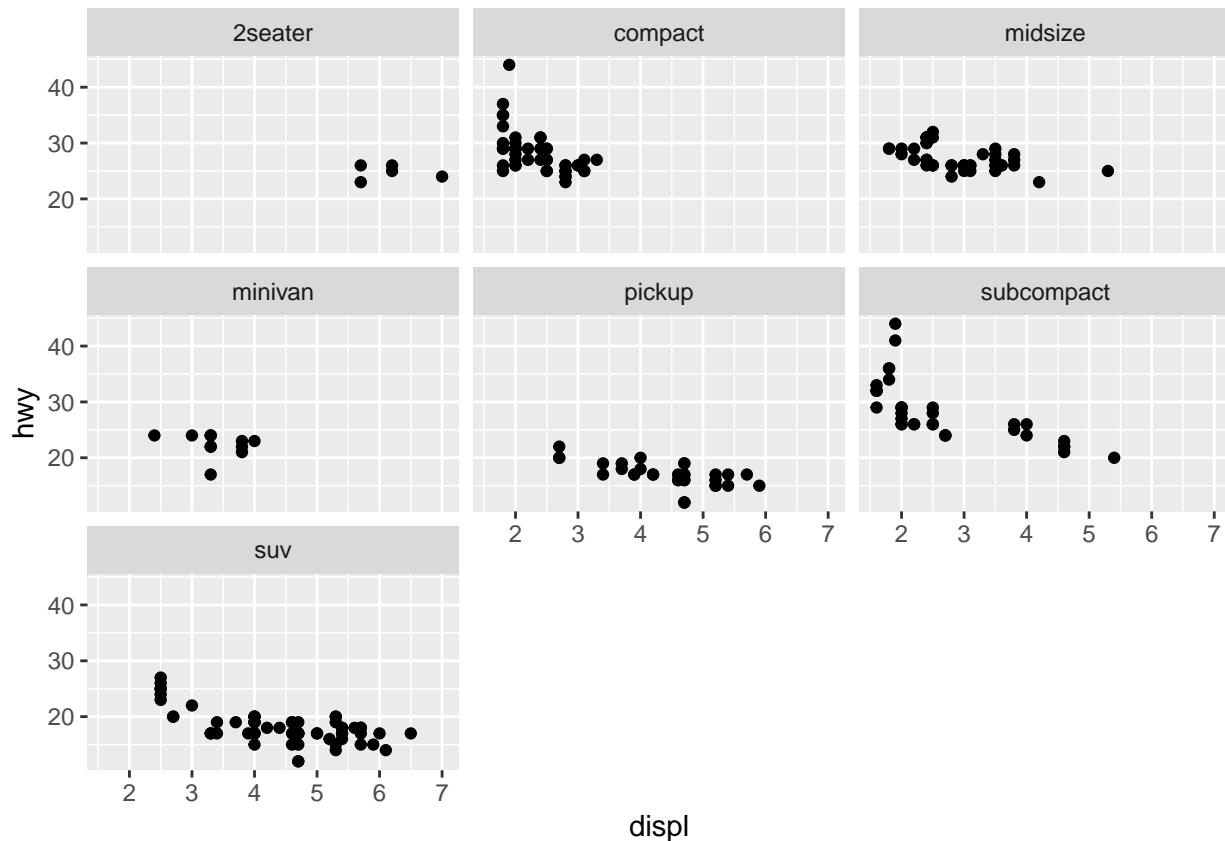
```
ggplot(mpg, aes(displ, hwy)) +  
  geom_point()
```



```
ggplot(mpg, aes(displ, hwy)) +  
  geom_point() +  
  NULL
```

```
ggplot(mpg, aes(displ, hwy)) +  
  geom_point() +  
  facet_wrap(~ class)
```



```
summary(lm(hwy ~ displ, mpg))
```

```
##
## Call:
## lm(formula = hwy ~ displ, data = mpg)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1039 -2.1646 -0.2242  2.0589 15.0105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.6977     0.7204   49.55  <2e-16 ***
## displ       -3.5306     0.1945  -18.15  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.836 on 232 degrees of freedom
## Multiple R-squared:  0.5868, Adjusted R-squared:  0.585
## F-statistic: 329.5 on 1 and 232 DF, p-value: < 2.2e-16
```

```
names(summary(lm(hwy ~ displ, mpg)))
```

```
## [1] "call"          "terms"         "residuals"    "coefficients"
## [5] "aliases"      "sigma"         "df"           "r.squared"
## [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
```

```
fits <- mpg %>%
  group_by(class) %>%
  do(tibble(fit = list(lm(hwy ~ displ, .))))
```

```
fits
```

```
## # A tibble: 7 x 2
## # Groups:   class [7]
##   class      fit
##   <chr>      <list>
## 1 2seater    <S3: lm>
## 2 compact   <S3: lm>
## 3 midsize   <S3: lm>
## 4 minivan   <S3: lm>
## 5 pickup    <S3: lm>
## 6 subcompact <S3: lm>
## 7 suv       <S3: lm>
```

```
fits %>%
  mutate(adjR2 = summary(fit[[1]])$adj.r.squared)
```

```
## # A tibble: 7 x 3
## # Groups:   class [7]
##   class      fit      adjR2
##   <chr>      <list>    <dbl>
## 1 2seater    <S3: lm> -0.313
## 2 compact   <S3: lm>  0.291
## 3 midsize   <S3: lm>  0.251
## 4 minivan   <S3: lm> -0.0568
## 5 pickup    <S3: lm>  0.370
## 6 subcompact <S3: lm>  0.461
## 7 suv       <S3: lm>  0.417
```

```
fits %>%
  rowwise() %>%
  mutate(adjR2 = summary(fit)$adj.r.squared)
```

```
## Source: local data frame [7 x 3]
## Groups: <by row>
##
## # A tibble: 7 x 3
##   class      fit      adjR2
##   <chr>      <list>    <dbl>
## 1 2seater    <S3: lm> -0.313
## 2 compact   <S3: lm>  0.291
## 3 midsize   <S3: lm>  0.251
## 4 minivan   <S3: lm> -0.0568
## 5 pickup    <S3: lm>  0.370
## 6 subcompact <S3: lm>  0.461
## 7 suv       <S3: lm>  0.417
```

Output

```
mpg %>%  
  group_by(class) %>%  
  summarise(mean_hwy = mean(hwy))
```

```
## # 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
```

```
mpg %>%  
  group_by(class) %>%  
  summarise(mean_hwy = mean(hwy)) %>%  
  pander::pander()
```

class	mean_hwy
2seater	24.8
compact	28.3
midsize	27.29
minivan	22.36
pickup	16.88
subcompact	28.14
suv	18.13

```
library(pander)
```

```
mpg %>%  
  group_by(class) %>%  
  summarise(mean_hwy = mean(hwy)) %>%  
  pander(digits = 3)
```

class	mean_hwy
2seater	24.8
compact	28.3
midsize	27.3
minivan	22.4
pickup	16.9
subcompact	28.1
suv	18.1