Untitled

Yunting Chiu

10/16/2020

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.3      v dplyr   1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(broom)
mpg %>%
  select_if(is.character)
```

```
## # A tibble: 234 x 6
##    manufacturer model      trans      drv   fl    class
##    <chr>        <chr>      <chr>      <chr> <chr> <chr>
##  1 audi         a4         auto(l5)   f     p     compact
##  2 audi         a4         manual(m5) f     p     compact
##  3 audi         a4         manual(m6) f     p     compact
##  4 audi         a4         auto(av)   f     p     compact
##  5 audi         a4         auto(l5)   f     p     compact
##  6 audi         a4         manual(m5) f     p     compact
##  7 audi         a4         auto(av)   f     p     compact
##  8 audi         a4 quattro manual(m5) 4     p     compact
##  9 audi         a4 quattro auto(l5)   4     p     compact
## 10 audi         a4 quattro manual(m6) 4     p     compact
## # ... with 224 more rows
```

```r
mtcars
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
```

```
## Merc 280C            17.8   6 167.6 123 3.92 3.440 18.90  1 0    4    4
## Merc 450SE           16.4   8 275.8 180 3.07 4.070 17.40  0 0    3    3
## Merc 450SL           17.3   8 275.8 180 3.07 3.730 17.60  0 0    3    3
## Merc 450SLC          15.2   8 275.8 180 3.07 3.780 18.00  0 0    3    3
## Cadillac Fleetwood   10.4   8 472.0 205 2.93 5.250 17.98  0 0    3    4
## Lincoln Continental  10.4   8 460.0 215 3.00 5.424 17.82  0 0    3    4
## Chrysler Imperial    14.7   8 440.0 230 3.23 5.345 17.42  0 0    3    4
## Fiat 128             32.4   4  78.7  66 4.08 2.200 19.47  1 1    4    1
## Honda Civic          30.4   4  75.7  52 4.93 1.615 18.52  1 1    4    2
## Toyota Corolla       33.9   4  71.1  65 4.22 1.835 19.90  1 1    4    1
## Toyota Corona        21.5   4 120.1  97 3.70 2.465 20.01  1 0    3    1
## Dodge Challenger     15.5   8 318.0 150 2.76 3.520 16.87  0 0    3    2
## AMC Javelin          15.2   8 304.0 150 3.15 3.435 17.30  0 0    3    2
## Camaro Z28           13.3   8 350.0 245 3.73 3.840 15.41  0 0    3    4
## Pontiac Firebird     19.2   8 400.0 175 3.08 3.845 17.05  0 0    3    2
## Fiat X1-9            27.3   4  79.0  66 4.08 1.935 18.90  1 1    4    1
## Porsche 914-2        26.0   4 120.3  91 4.43 2.140 16.70  0 1    5    2
## Lotus Europa         30.4   4  95.1 113 3.77 1.513 16.90  1 1    5    2
## Ford Pantera L       15.8   8 351.0 264 4.22 3.170 14.50  0 1    5    4
## Ferrari Dino         19.7   6 145.0 175 3.62 2.770 15.50  0 1    5    6
## Maserati Bora        15.0   8 301.0 335 3.54 3.570 14.60  0 1    5    8
## Volvo 142E           21.4   4 121.0 109 4.11 2.780 18.60  1 1    4    2
```

- Transform the data so AC, Pool and Highway are factors and Price is in thousands of dollars.

```r
# Transform the data so AC, Pool and Highway are factors and Price is in thousands of dollars.
estate <- read_csv("../housing_app/data/estate.csv",
             col_types = cols("AC" = col_factor(),
                              "Pool" = col_factor(),
                              "Highway" = col_factor())) %>%
       mutate(Price = Price/1000) %>%
       rename("Price($K)" = "Price") %>%
       mutate(AC = fct_recode(AC, "Presence" = "1", "Absence" = "0"),
              Pool = fct_recode(Pool, "Pool" = "1", "No Pool" = "0"),
              Highway = fct_recode(Highway, "Adjacent" = "1", "Not Adjacent" = "0")) -> estate
estate
```

```
## # A tibble: 522 x 12
##    `Price($K)`  Area   Bed  Bath AC      Garage Pool    Year Quality Style   Lot
##          <dbl> <dbl> <dbl> <dbl> <fct>    <dbl> <fct>  <dbl> <chr>   <dbl> <dbl>
##  1         360  3032     4     4 Pres~        2 No P~   1972 Medium      1 22221
##  2         340  2058     4     2 Pres~        2 No P~   1976 Medium      1 22912
##  3         250  1780     4     3 Pres~        2 No P~   1980 Medium      1 21345
##  4         206. 1638     4     2 Pres~        2 No P~   1963 Medium      1 17342
##  5         276. 2196     4     3 Pres~        2 No P~   1968 Medium      7 21786
##  6         248  1966     4     3 Pres~        5 Pool   1972 Medium      1 18902
##  7         230. 2216     3     2 Pres~        2 No P~   1972 Medium      7 18639
##  8         150  1597     2     1 Pres~        1 No P~   1955 Medium      1 22112
##  9         195  1622     3     2 Pres~        2 No P~   1975 Low         1 14321
## 10         160  1976     3     3 Abse~        1 No P~   1918 Low         1 32358
## # ... with 512 more rows, and 1 more variable: Highway <fct>
```

```r
estate %>%
  select(Area,Year) %>%
    log() -> RegData
augment(lm(Area~Year, data = RegData)) %>%
```

```r
ggplot(aes(x = .fitted, y = .resid)) + geom_point()
```
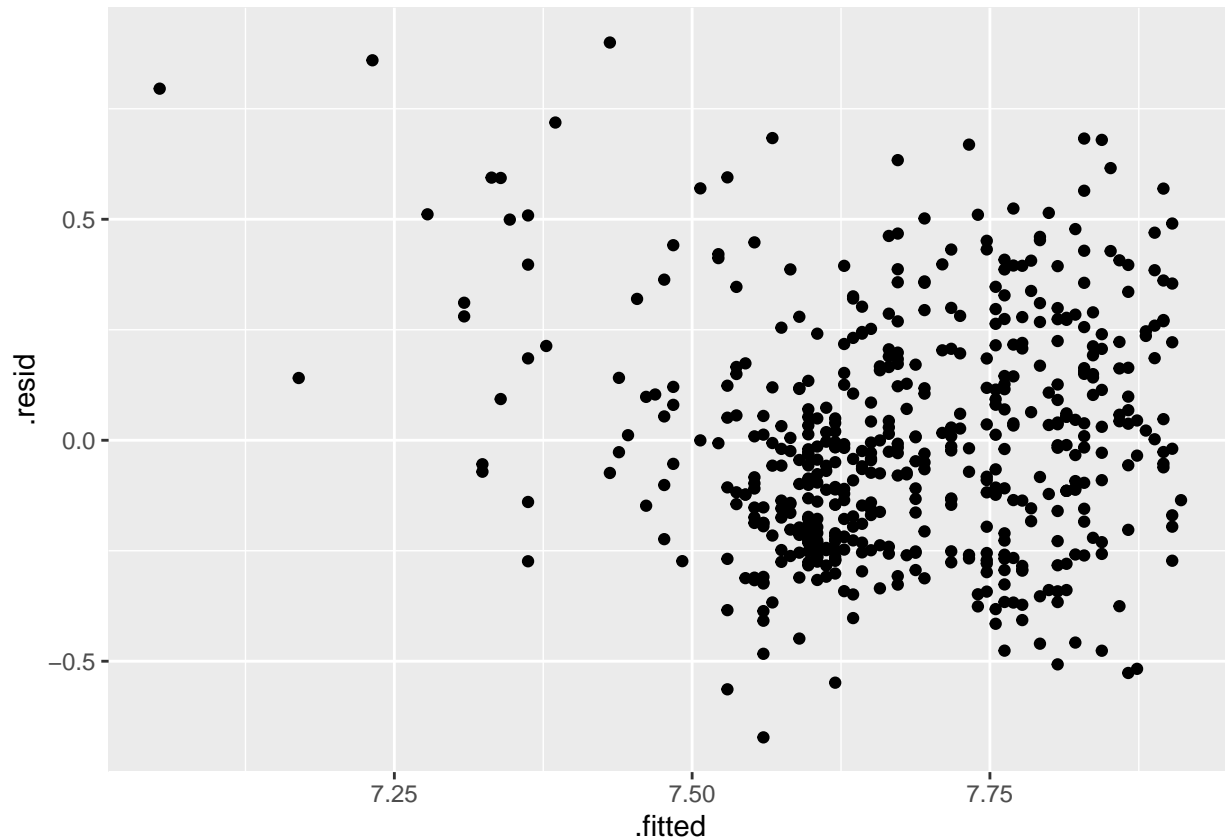

- check NA

```r
estate %>%
summarize(across(everything(), ~sum(is.na(.))))
```

```
## # A tibble: 1 x 12
##   `Price($K)`  Area   Bed  Bath    AC Garage  Pool  Year Quality Style   Lot
##         <int> <int> <int> <int> <int>  <int> <int> <int>   <int> <int> <int>
## 1           0     0     0     0     0      0     0     0       0     0     0
## # ... with 1 more variable: Highway <int>
```
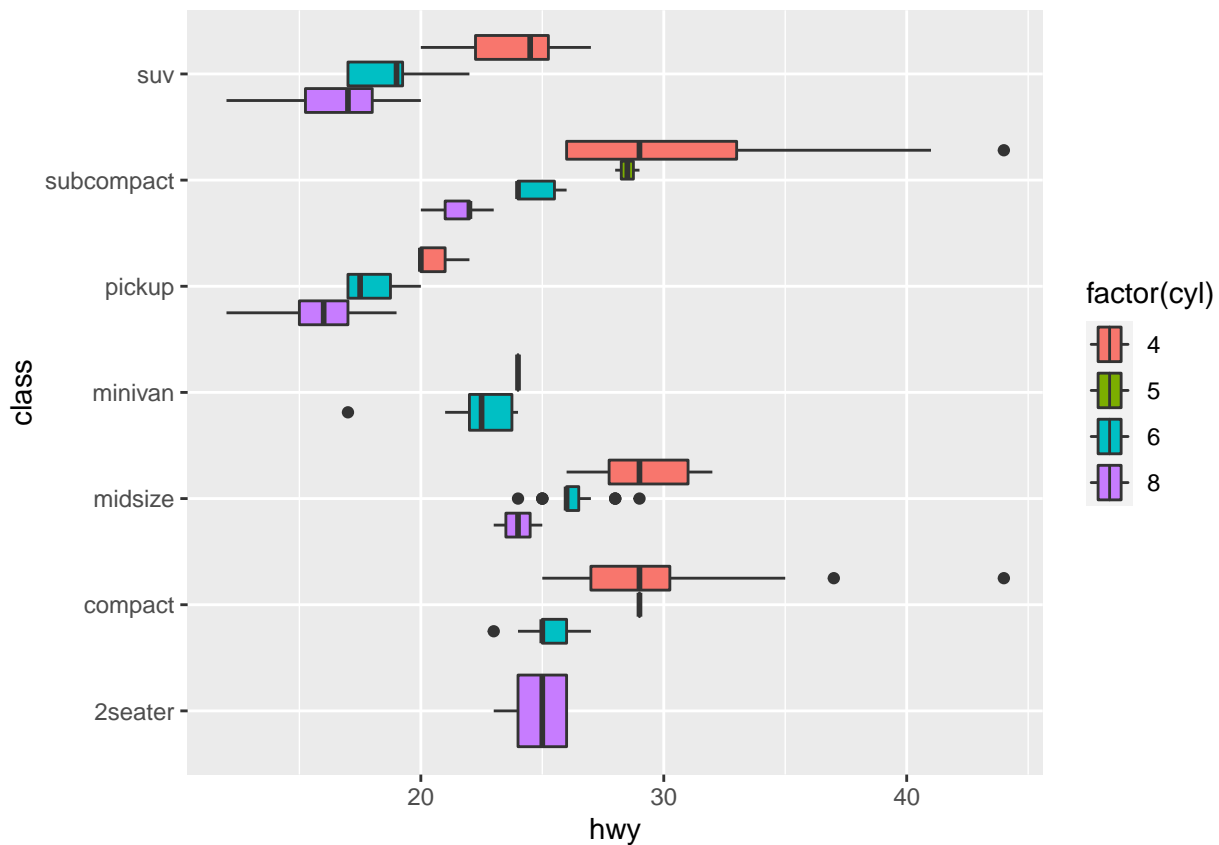
```r
library(ggstance)
```

```
##
## Attaching package: 'ggstance'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
```

```r
# With ggplot2 we need coord_flip():
ggplot(mpg, aes(class, hwy, fill = factor(cyl))) +
  geom_boxplot() +
  coord_flip()
```
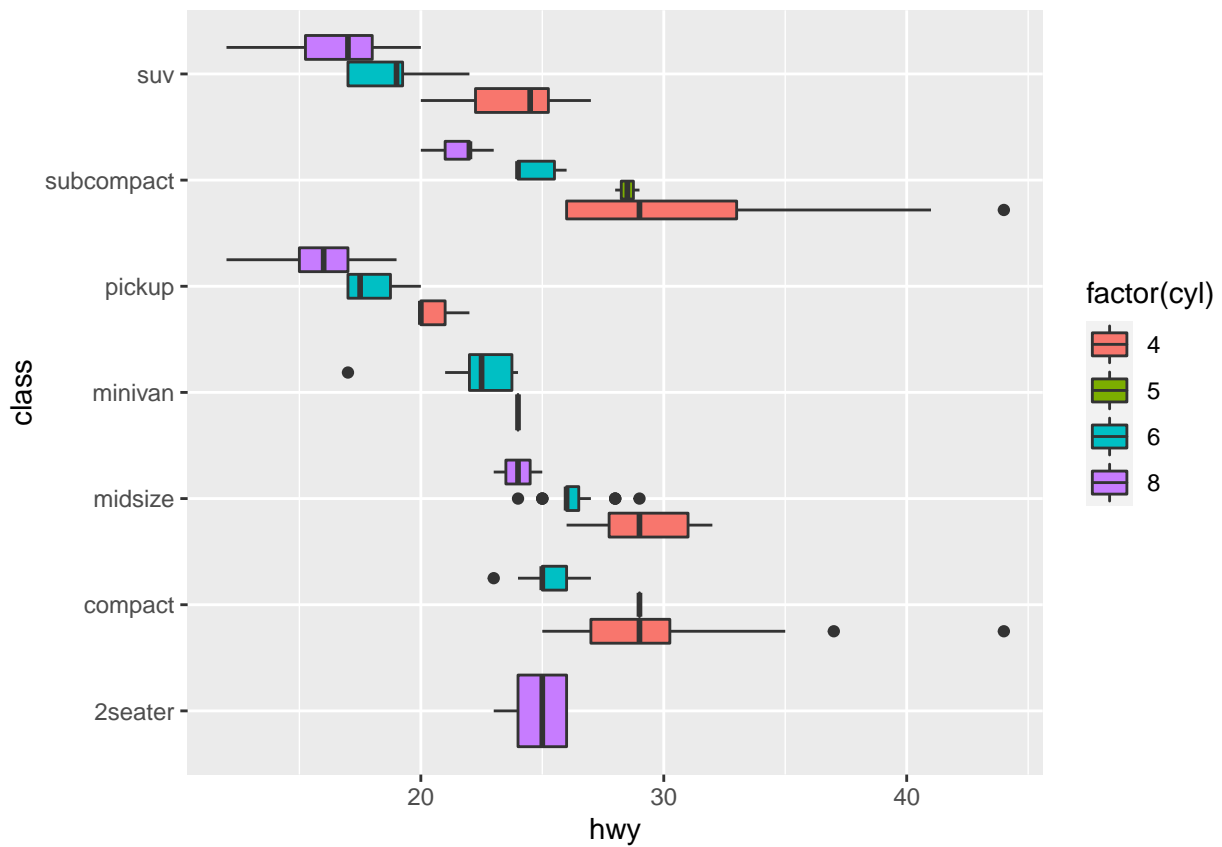
```
# With ggstance we use the h-suffixed version:
ggplot(mpg, aes(hwy, class, fill = factor(cyl))) +
  geom_boxploth()
```

```r
# With facets ggstance horizontal layers are often the only way of
# having all ggplot features working correctly, for instance free
# scales:
df <- data.frame(
  Group = factor(rep(1:3, each = 4), labels = c("Drug A", "Drug B", "Control")),
  Subject = factor(rep(1:6, each = 2), labels = c("A", "B", "C", "D", "E", "F")),
  Result = rnorm(12)
)

ggplot(df, aes(Result, Subject))+
  geom_boxploth(aes(fill = Group))+
  facet_grid(Group ~ ., scales = "free_y")
```