# Panel plots in R

Joyce Robbins

Agenda

- Basic faceting in `ggplot2`

- When to free scales

- Getting data in the right form

- Plotting one variable per panel

- Design decisions with multiple variables

- Scatterplot matrices

Slides: www.github.com/jtr13/PanelPlots/NYCASAMetro.pdf

Code: www.github.com/jtr13/PanelPlots/NYCASAMetro.Rmd
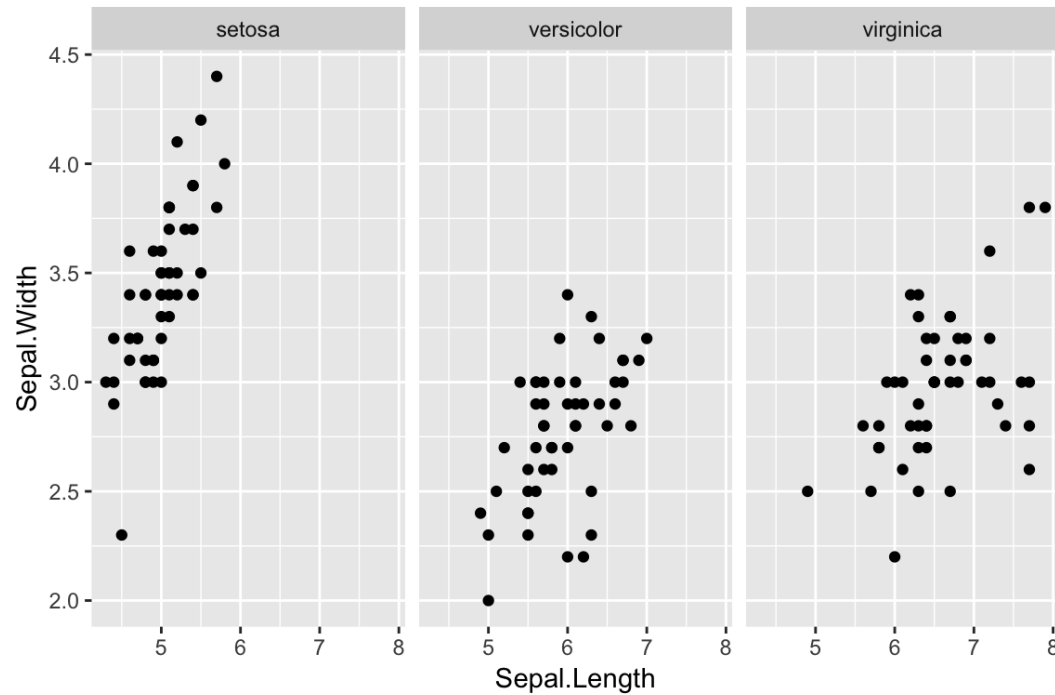
Without faceting

```r
library(tidyverse)
g <- ggplot(iris, aes(Sepal.Length, Sepal.Width)) + geom_point()
g
```
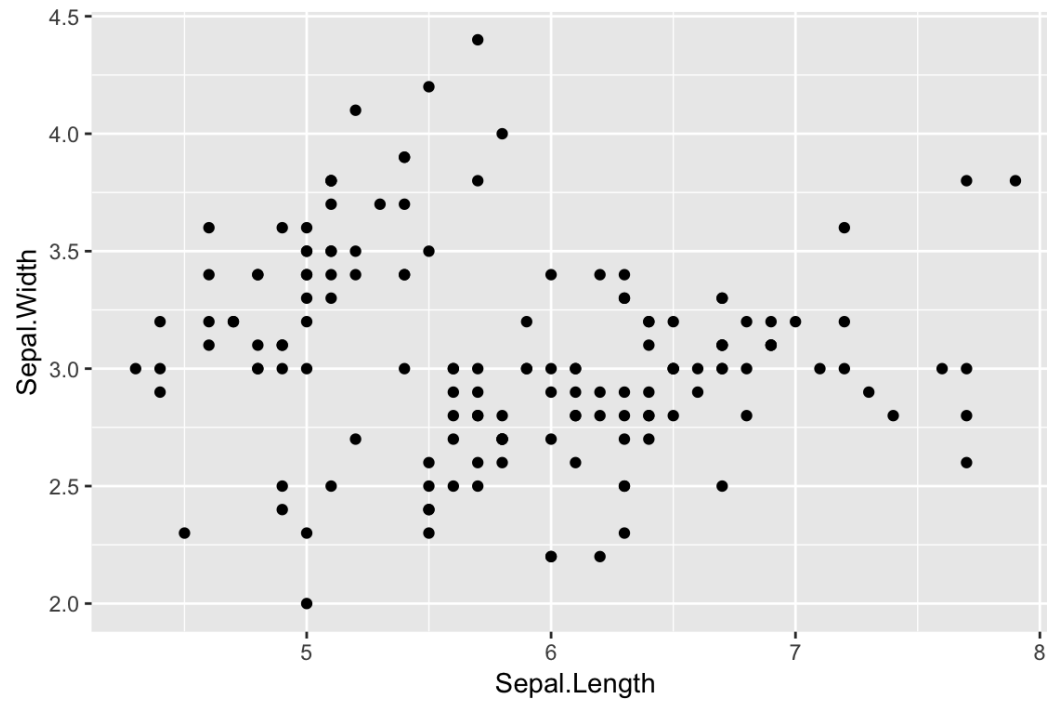
Facet on one variable with `facet_wrap()`

facet "on" Species

```
g + facet_wrap(~Species)
```

Faceting in ggplot2

Each panel represents one categorical group / levels of a factor (type can be factor or character or integer)

```r
glimpse(iris)
```

```
## Observations: 150
## Variables: 5
## $ Sepal.Length <dbl> 5.1, 4.9, 4.7, 4.6, 5.0, 5.4, 4.6, 5.0, 4.4, 4.9, 5…
## $ Sepal.Width  <dbl> 3.5, 3.0, 3.2, 3.1, 3.6, 3.9, 3.4, 3.4, 2.9, 3.1, 3…
## $ Petal.Length <dbl> 1.4, 1.4, 1.3, 1.5, 1.4, 1.7, 1.4, 1.5, 1.4, 1.5, 1…
## $ Petal.Width  <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0…
## $ Species      <fct> setosa, setosa, setosa, setosa, setosa, setosa, set…
```
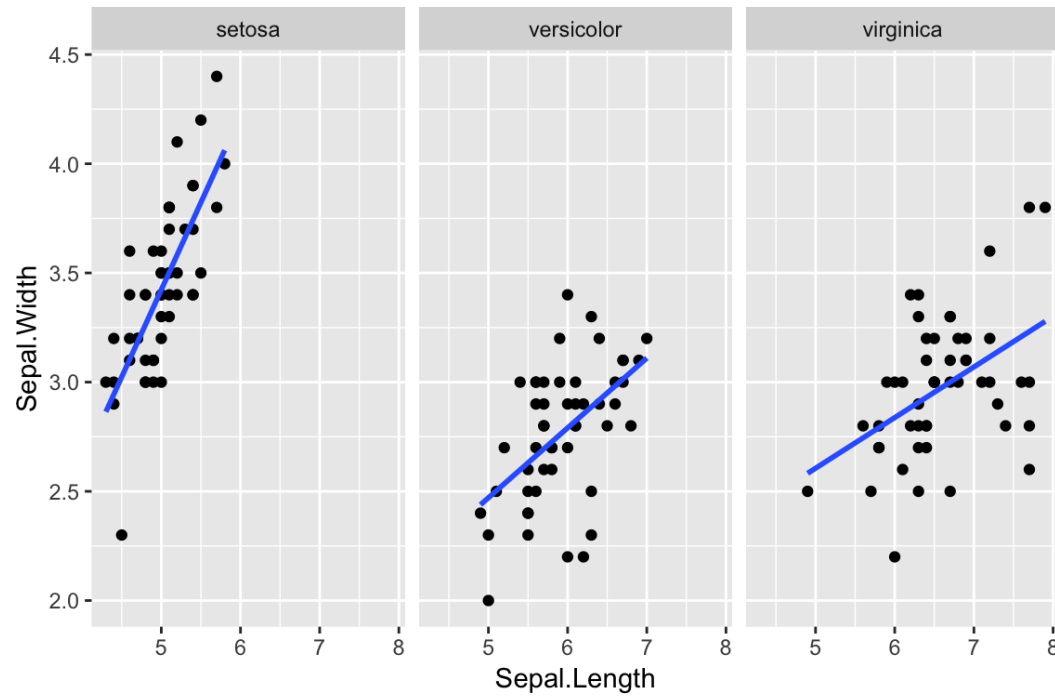
Change the layout with `nrow, ncol`

```
g + facet_wrap(~Species, ncol = 1)
```
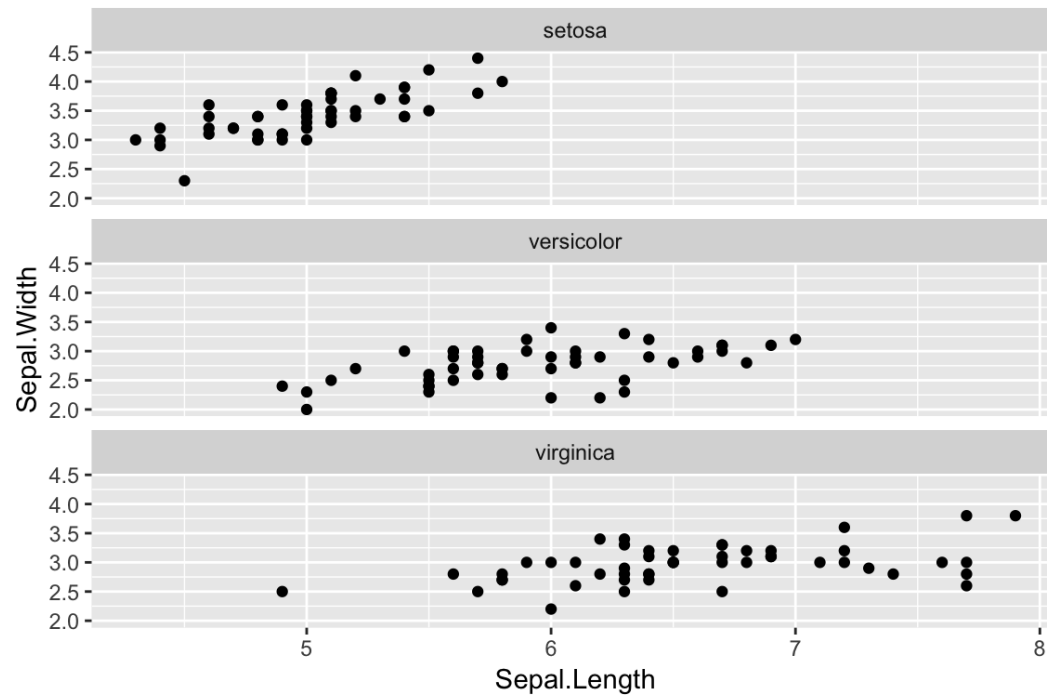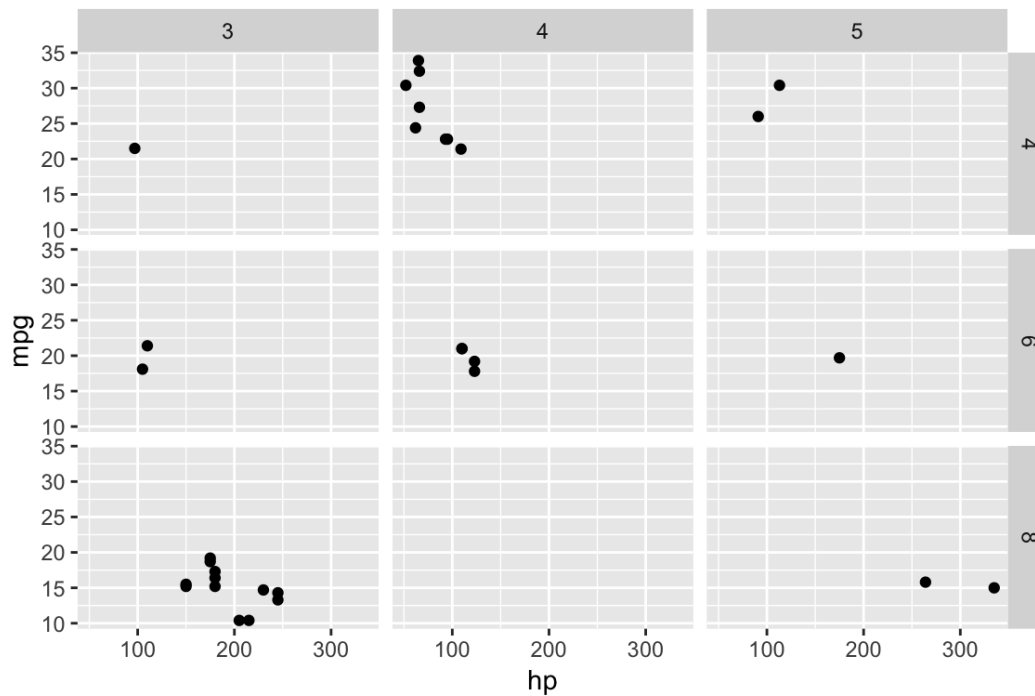
Add regression lines

```
g + geom_smooth(method = "lm", se = FALSE) + facet_wrap(~Species)
```
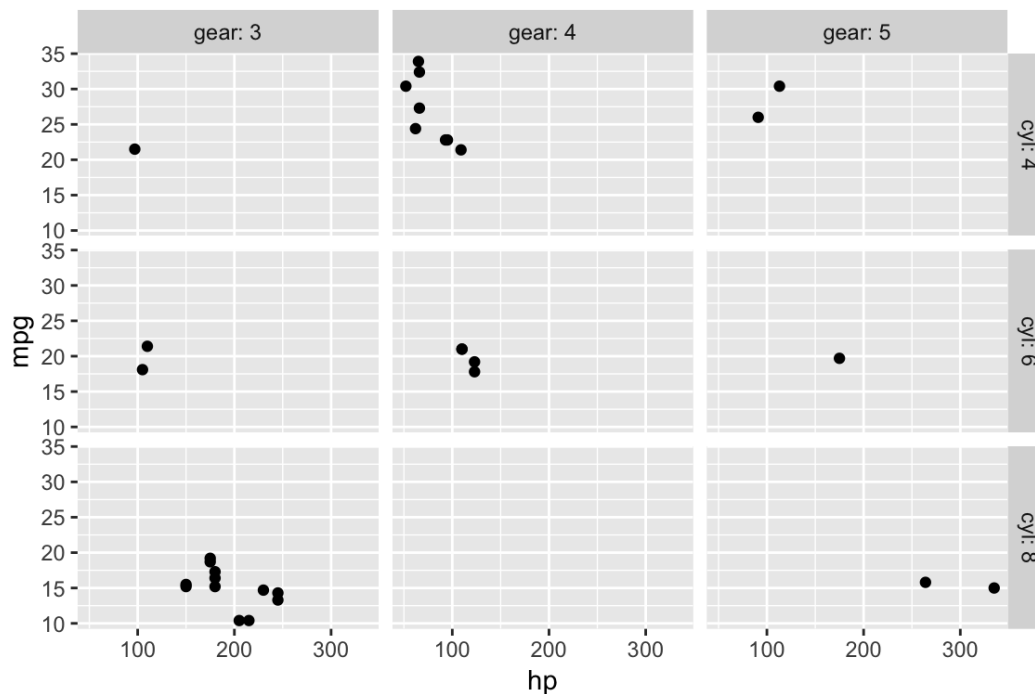
Facet on two (categorical) variables with `facet_grid()`

```
ggplot(mtcars, aes(hp,mpg)) + geom_point() + facet_grid(cyl~gear)
```

Label variables (in addition to factor levels)

```
ggplot(mtcars, aes(hp, mpg)) + geom_point() + facet_grid(cyl~gear, labeller = label_both)
```
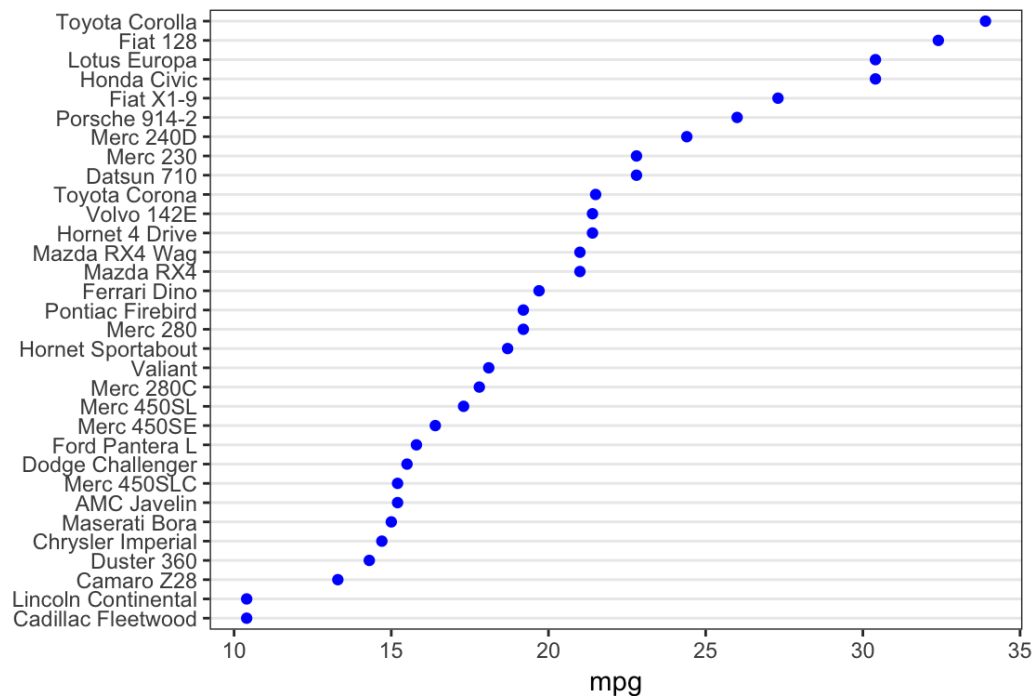
Cleveland dot plots
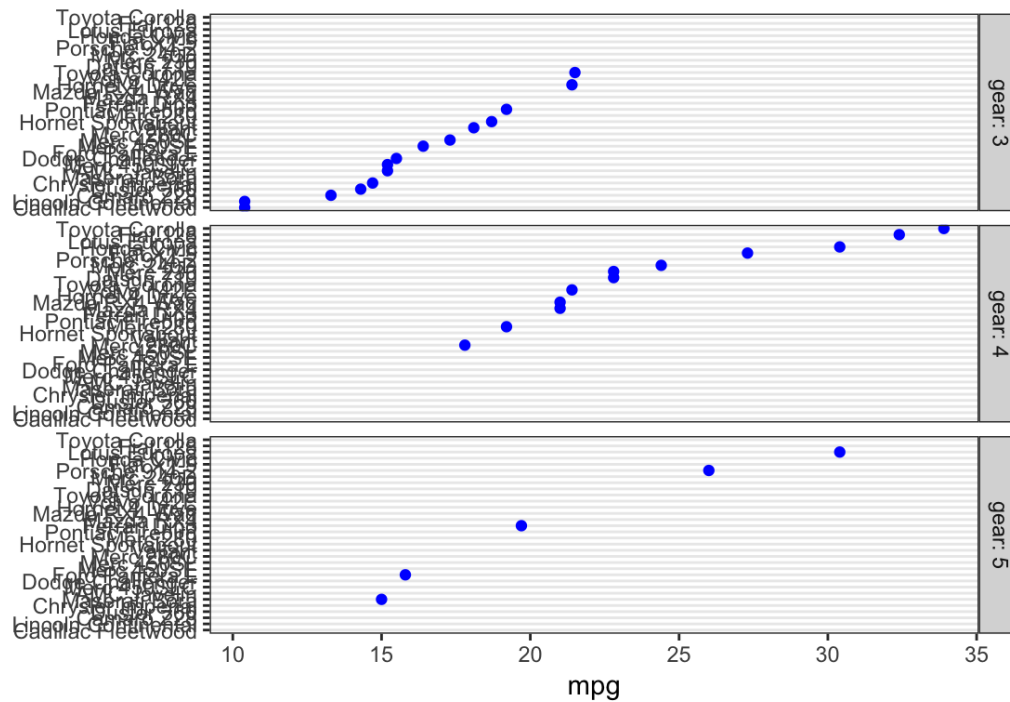
Note that y-axis is discrete

```
g <- mtcars %>% rownames_to_column("car") %>%
  ggplot(aes(mpg, reorder(car, mpg))) + geom_point(color = "blue") + theme_bw() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank()) + ylab("")
g
```
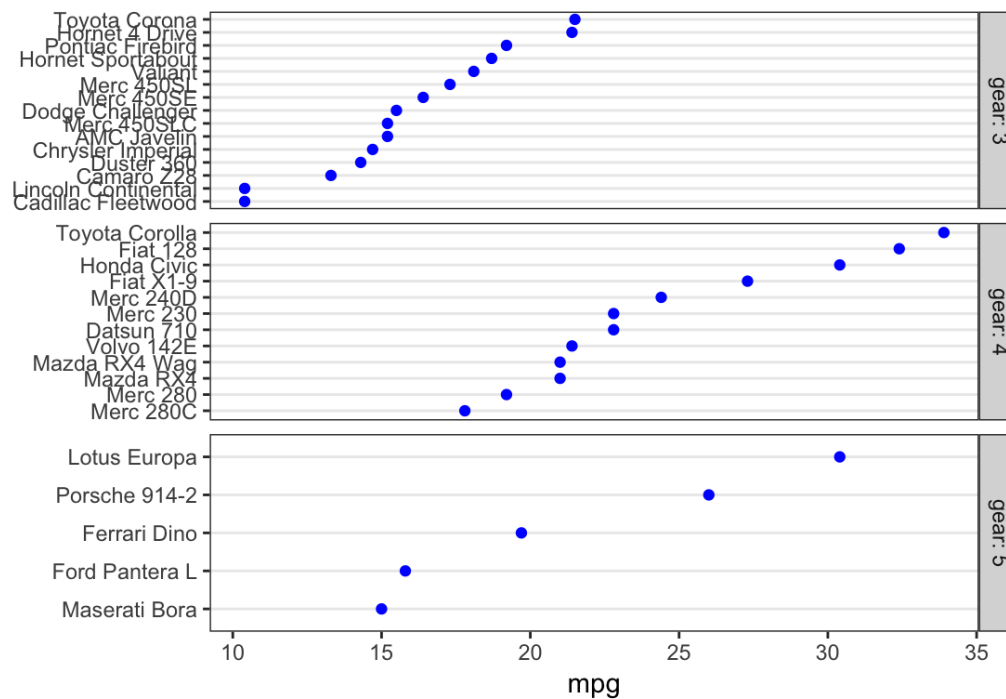
Facet by gear

```
g + facet_grid(gear ~ .,
               labeller = label_both)
```

"Free" the y scale with `scales = "free_y"`

```
g + facet_grid(gear ~ ., labeller = label_both, scales = "free_y")
```

Change panel heights with
`space = "free_y"`

```
g + facet_grid(gear ~ ., labeller = label_both, scales = "free_y",
               space = "free_y")
```
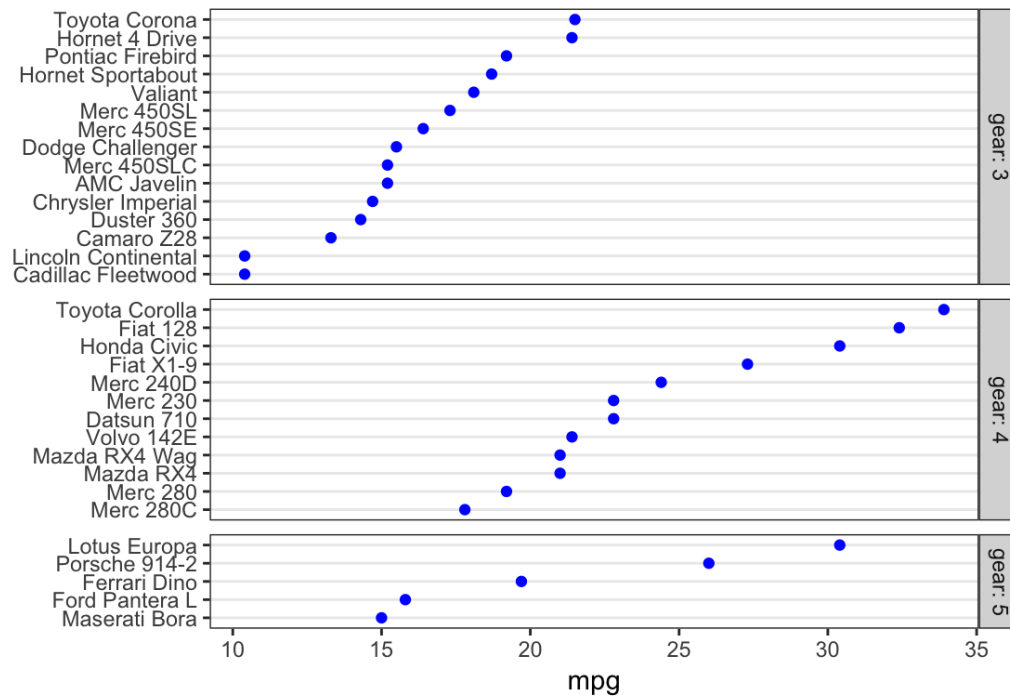
In general, do not "free" numerical scales

```
ggplot(iris, aes(Sepal.Length, Sepal.Width)) +
  geom_point() + facet_wrap(~Species, scales = "free")
```
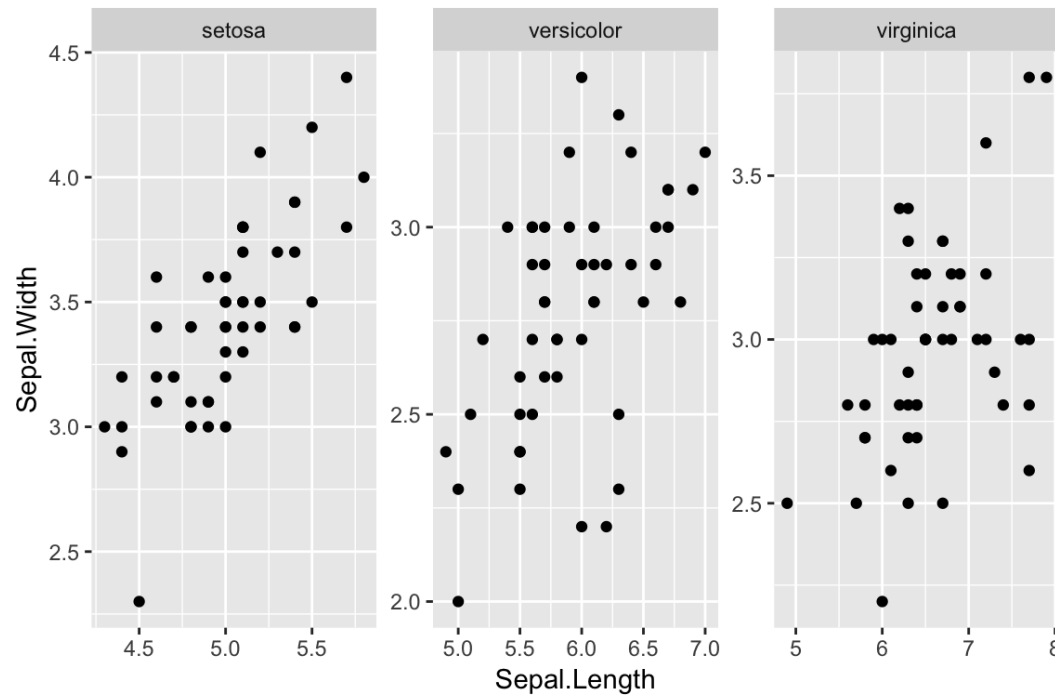

incorrect

In general, do not "free" numerical scales

```r
ggplot(iris, aes(Sepal.Length, Sepal.Width)) +
   geom_point() + facet_wrap(~Species)
```


correct

Data must be in the right form to facet

One column of categorical data = one faceting direction

```
library(ggplot2movies)
str(movies)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':     58788 obs. of   24 variables:
##  $ title     : chr  "$" "$1000 a Touchdown" "$21 a Day Once a Month" "$40,000" ...
##  $ year      : int  1971 1939 1941 1996 1975 2000 2002 2002 1987 1917 ...
##  $ length    : int  121 71 7 70 71 91 93 25 97 61 ...
##  $ budget    : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ rating    : num  6.4 6 8.2 8.2 3.4 4.3 5.3 6.7 6.6 6 ...
##  $ votes     : int  348 20 5 6 17 45 200 24 18 51 ...
##  $ r1        : num  4.5 0 0 14.5 24.5 4.5 4.5 4.5 4.5 4.5 ...
##  $ r2        : num  4.5 14.5 0 0 4.5 4.5 0 4.5 4.5 0 ...
##  $ r3        : num  4.5 4.5 0 0 0 4.5 4.5 4.5 4.5 4.5 ...
##  $ r4        : num  4.5 24.5 0 0 14.5 14.5 4.5 4.5 0 4.5 ...
##  $ r5        : num  14.5 14.5 0 0 14.5 14.5 24.5 4.5 0 4.5 ...
##  $ r6        : num  24.5 14.5 24.5 0 4.5 14.5 24.5 14.5 0 44.5 ...
##  $ r7        : num  24.5 14.5 0 0 0 4.5 14.5 14.5 34.5 14.5 ...
##  $ r8        : num  14.5 4.5 44.5 0 0 4.5 4.5 14.5 14.5 4.5 ...
##  $ r9        : num  4.5 4.5 24.5 34.5 0 14.5 4.5 4.5 4.5 4.5 ...
##  $ r10       : num  4.5 14.5 24.5 45.5 24.5 14.5 14.5 14.5 24.5 4.5 ...
##  $ mpaa      : chr  "" "" "" "" ...
##  $ Action    : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ Animation : int  0 0 1 0 0 0 0 0 0 0 ...
```
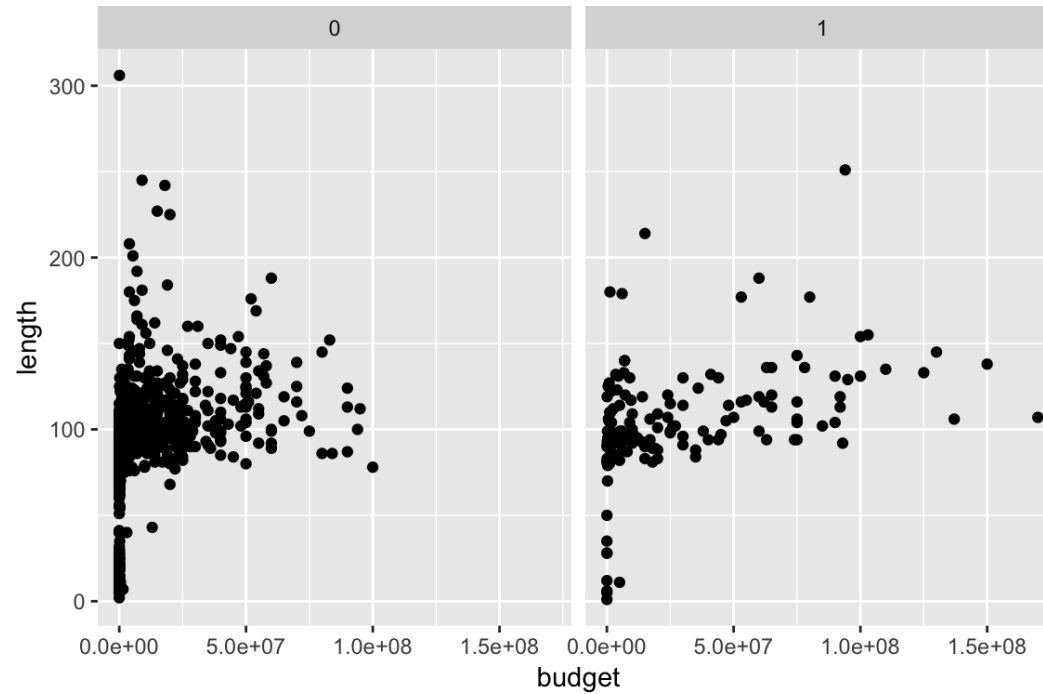
```
##   $ Comedy     : int  1 1 0 1 0 0 0 0 0 0 ...
##   $ Drama      : int  1 0 0 0 0 1 1 0 1 0 ...
##   $ Documentary: int  0 0 0 0 0 0 0 1 0 0 ...
##   $ Romance    : int  0 0 0 0 0 0 0 0 0 0 ...
##   $ Short      : int  0 0 1 0 0 0 0 1 0 0 ...
```

Facet on Action column

```
movies %>% filter(!is.na(budget)) %>% sample_n(1000) %>%
ggplot(aes(budget, length)) + geom_point() +
  facet_wrap(~Action)
```

Transform data

```
mymovies <- movies %>%
  select(title, length, budget, year, mpaa, Action:Short) %>%
  gather(key = "genre", value, Action:Short) %>%
  filter(value == 1) %>%
  select(-value)


head(mymovies)
```

```
## # A tibble: 6 x 6
##   title                length budget  year mpaa  genre
##   <chr>                 <int>  <int> <int> <chr> <chr>
## 1 $windle                  93     NA  2002 R     Action
## 2 'A' gai waak            106     NA  1983 PG-13 Action
## 3 'A' gai waak juk jaap   101     NA  1987 PG-13 Action
## 4 'Crocodile' Dundee II   110     NA  1988 ""    Action
## 5 'Gator Bait              88     NA  1974 ""    Action
## 6 'Sheba, Baby'            90     NA  1975 ""    Action
```

Facet on genre

```r
lengthorder <- mymovies %>% group_by(genre) %>% summarize(meanlength = mean(length)) %>%
  arrange(desc(meanlength))
mymovies %>% filter(!is.na(budget)) %>%
  sample_n(1000) %>%
ggplot(aes(budget/1000000, length)) + geom_point() +
  facet_wrap(~factor(genre, levels = lengthorder$genre)) + xlab("budget (in millions)") +
  geom_hline(data = lengthorder, aes(yintercept = meanlength), color = "red") +
  ggtitle ("Length vs. Budget", subtitle = "Red line indicates mean length")
```


Length vs. Budget
Red line indicates mean length
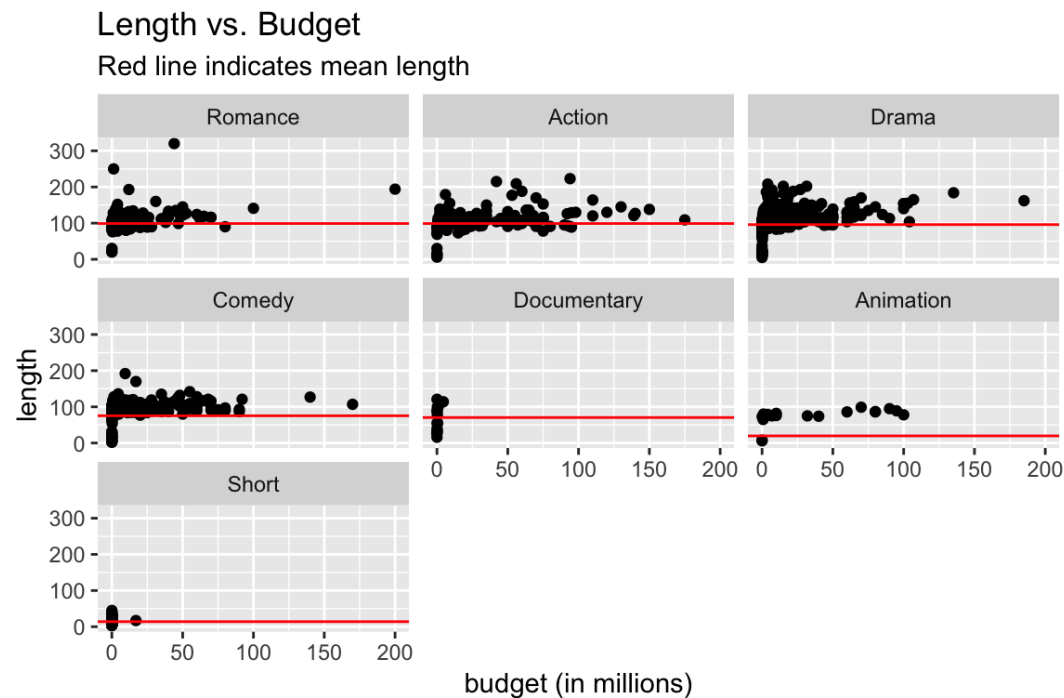
Use facets to look at distributions of numeric variables

```
newmovies <- movies %>% select(mpaa, year:votes) %>% gather(key = "variable", value, -mpaa) %>%
  filter(!is.na(value))
newmovies %>% sample_n(6)
```

```
## # A tibble: 6 x 3
##   mpaa  variable      value
##   <chr> <chr>         <dbl>
## 1 R     budget     20000000
## 2 ""    year           1985
## 3 ""    length           71
## 4 ""    year           2004
## 5 ""    votes            19
## 6 ""    rating          6.6
```
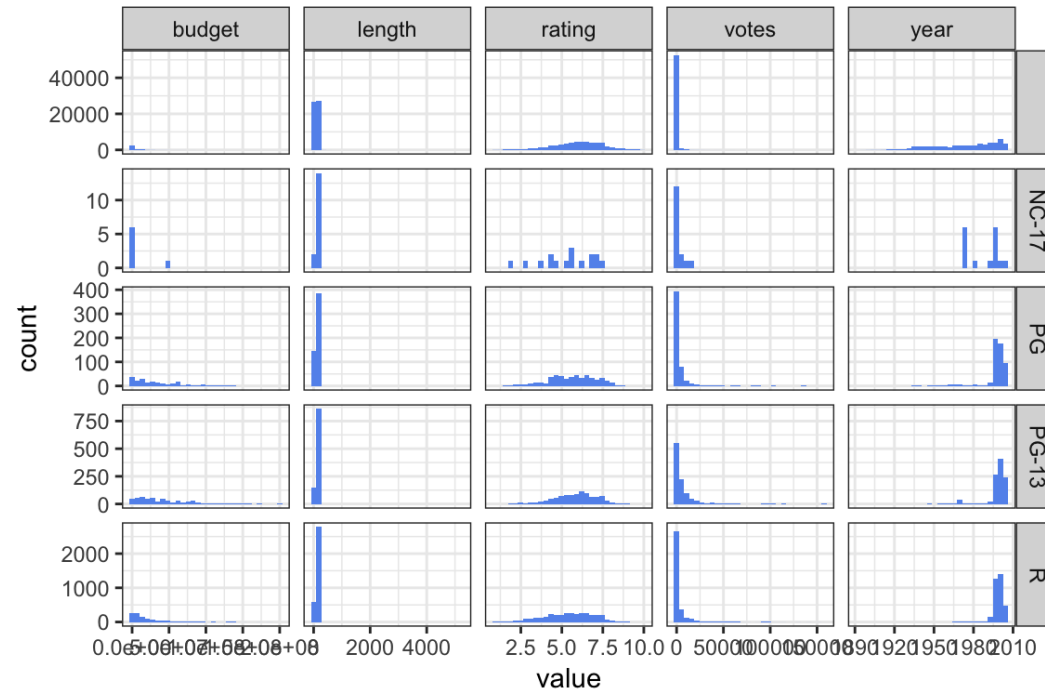
```
ggplot(newmovies, aes(value)) + geom_histogram(fill = "cornflowerblue") + facet_wrap(~variable, scales =
  "free")
```
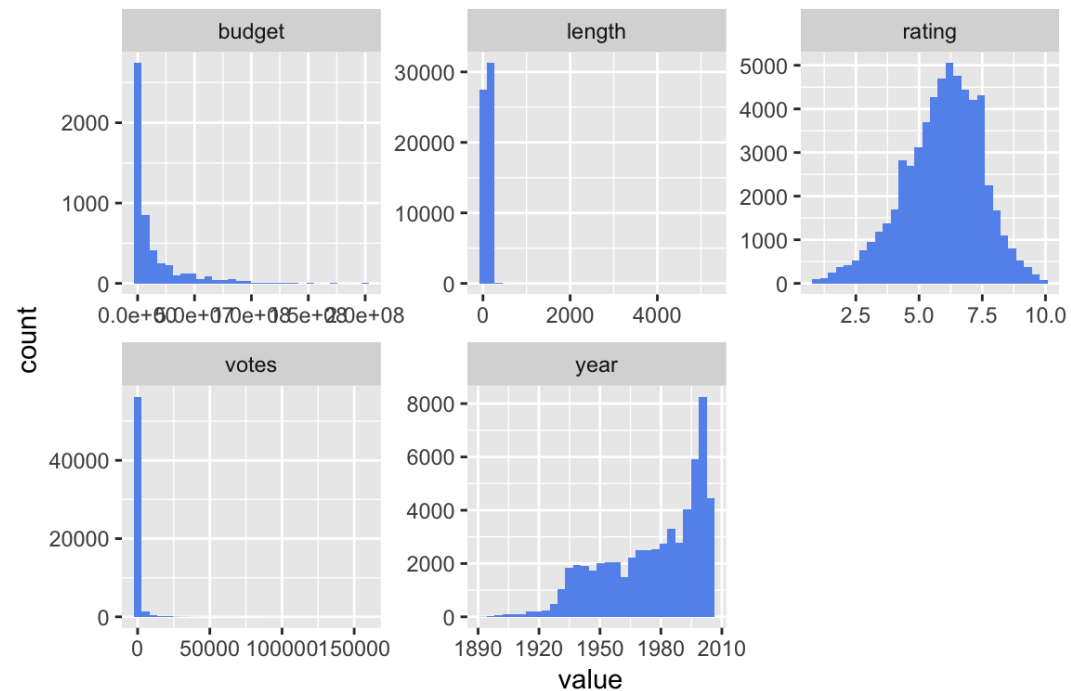

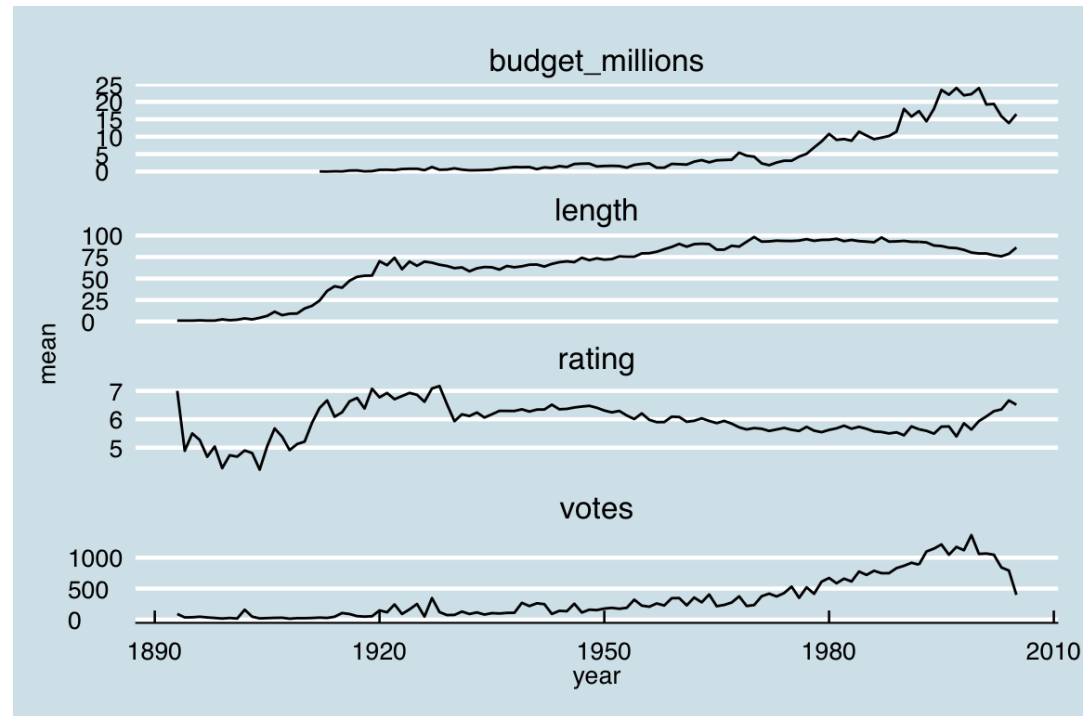variables become levels of the new "variable" column

Facet on `mpaa` and `variable`

```
ggplot(newmovies, aes(value)) + geom_histogram(fill = "cornflowerblue") + facet_grid(mpaa~variable,
    scales = "free") + theme_bw()
```

One common scale

Same technique: variables become levels of a new "variable" column

Data

```r
byyear <- movies %>% select(title:votes) %>%
  mutate(budget_millions = budget/1000000) %>%
  select(-budget) %>%
  gather(key = "variable", value, -title, -year) %>%
  group_by(year, variable) %>%
  summarize(mean = mean(value, na.rm = T))
```

```r
glimpse(byyear)
```

```
## Observations: 452
## Variables: 3
## Groups: year [113]
## $ year     <int> 1893, 1893, 1893, 1893, 1894, 1894, 1894, 1894, 1895, 1…
## $ variable <chr> "budget_millions", "length", "rating", "votes", "budget…
## $ mean     <dbl> NaN, 1.000000, 7.000000, 90.000000, NaN, 1.000000, 4.88…
```

```r
ggplot(byyear, aes(year, mean)) + geom_line() +
  facet_wrap(~variable, ncol = 1, scales = "free_y") + ggthemes::theme_economist()
```

# Convert numerical to categorical variables

```
mymovies <- mymovies %>% filter(length <= 180) %>%
  mutate(decade = factor(round(year/10)*10))


ggplot(mymovies, aes(length)) + geom_histogram(fill = "cornflowerblue") +
  facet_wrap(~decade)
```
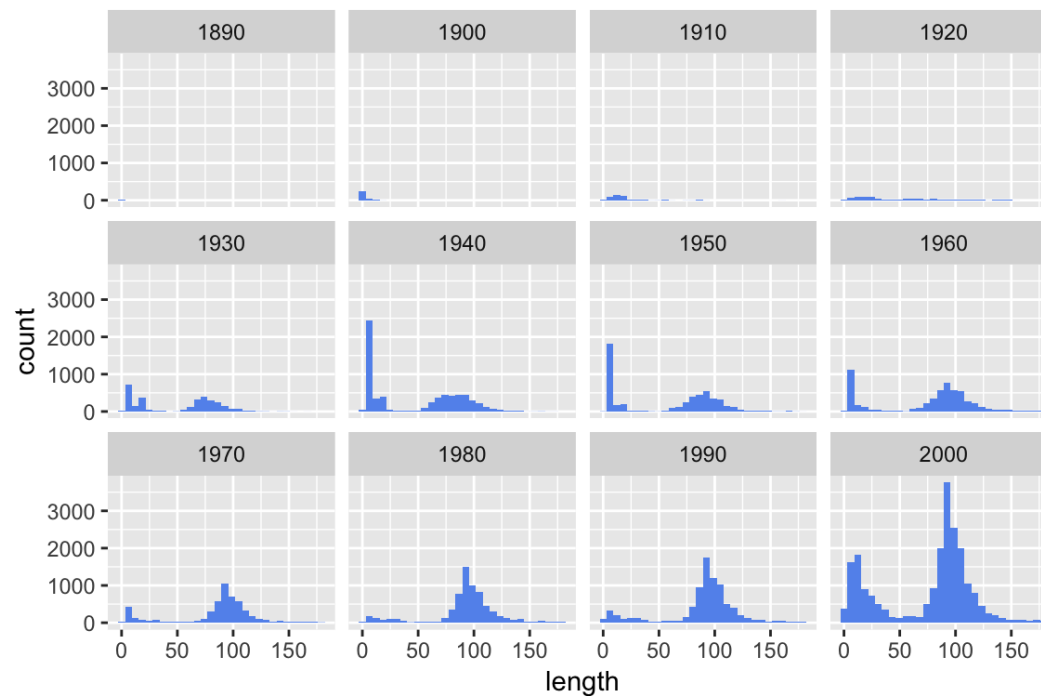
ggridges

```
library(ggridges)

ggplot(mymovies, aes(x=length, y=fct_rev(decade))) +
  geom_density_ridges(scale = 1.5, color = "blue", fill = "blue", alpha = .4) +
  xlab("Length (in minutes)") +
  ylab("") +
  theme_ridges()
```
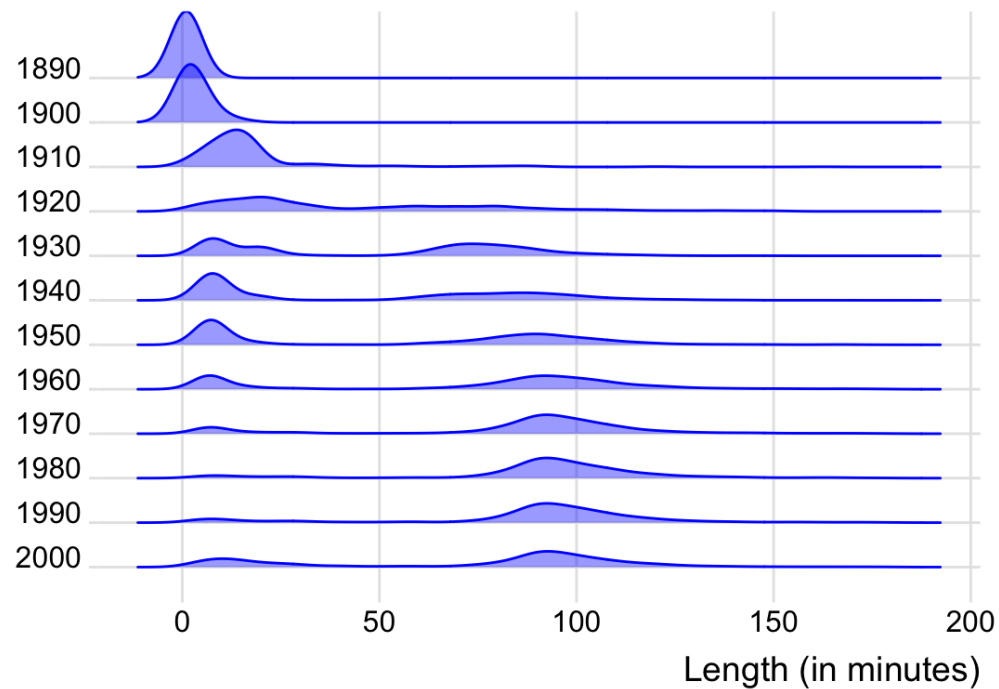
Design choices with multiple variables

x-axis, y-axis, row facets, column facets

color, size, shape

Think about continuous vs. categorical variables

```r
sales <- read_csv("sales2.csv")
glimpse(sales)
```

```
## Observations: 56,702
## Variables: 5
## $ `Order method type` <chr> "Telephone", "Telephone", "Telephone", "Tele…
## $ `Retailer type`     <chr> "Department", "Department", "Department", "D…
## $ `Product line`      <chr> "Camping", "Camping", "Camping", "Camping", …
## $ Revenue             <dbl> 0.01809251, 0.08225408, 0.02143473, 0.070400…
## $ Date                <date> 2012-03-30, 2012-03-30, 2012-03-30, 2012-03…
```
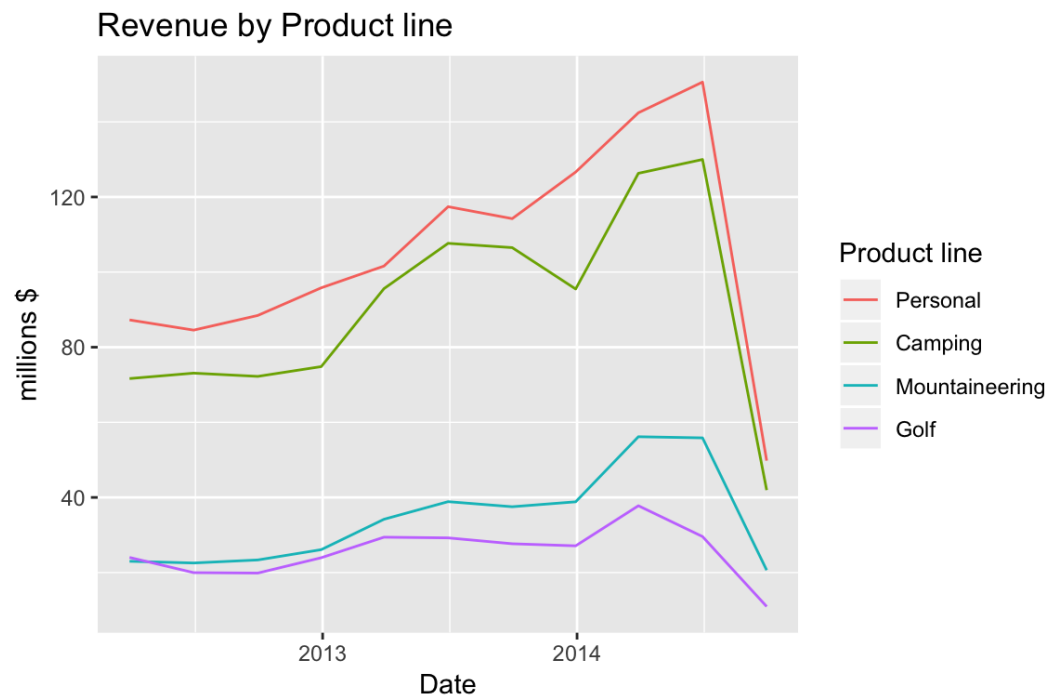
Start with one dimension

```
sales %>% group_by(Date) %>% summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev)) + geom_line() + ylab("millions $") + ggtitle("Revenue")
```
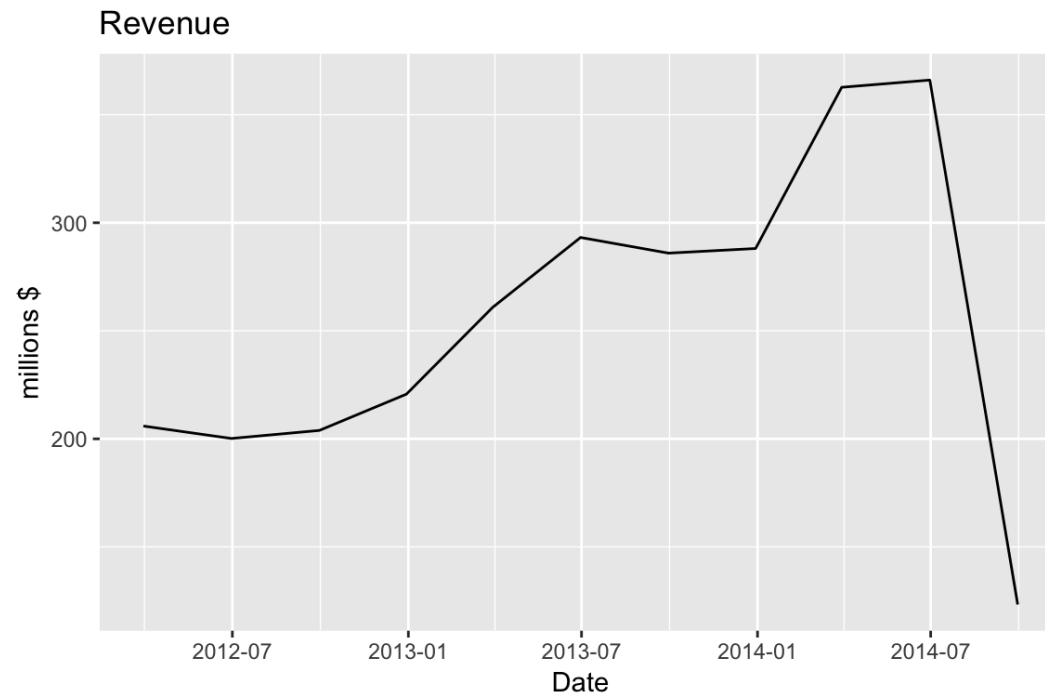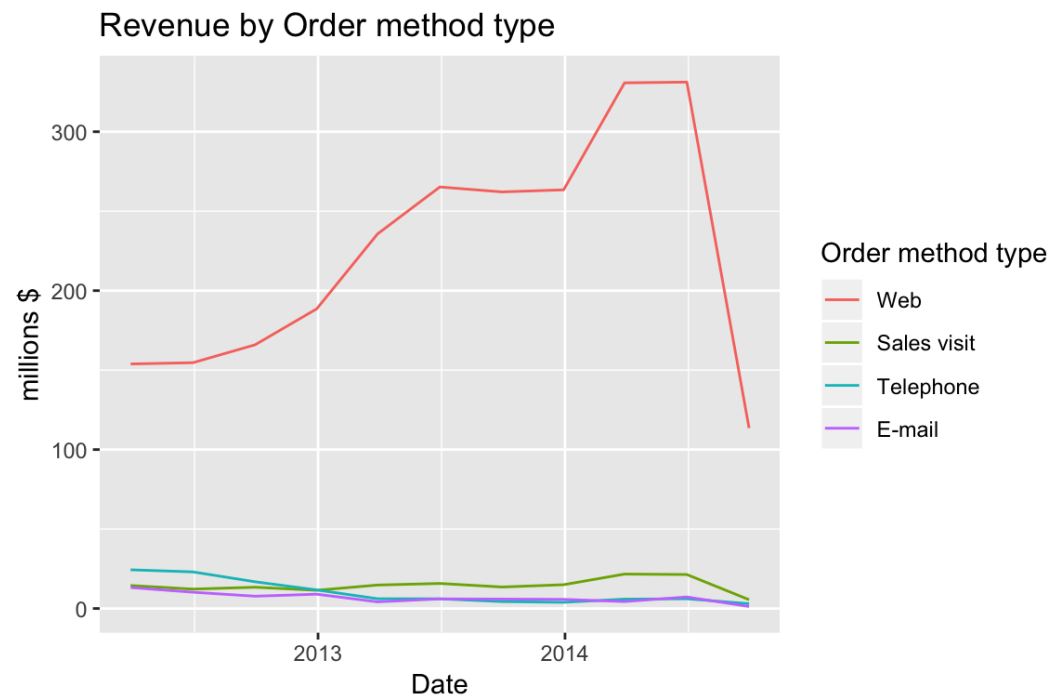
Consider one variable at a time

```
sales %>% group_by(Date, `Product line`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = fct_reorder2(`Product line`, Date, SumRev))) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year")  + ylab("millions $") + ggtitle("Revenue by
  Product line") +
  labs(color = "Product line")
```
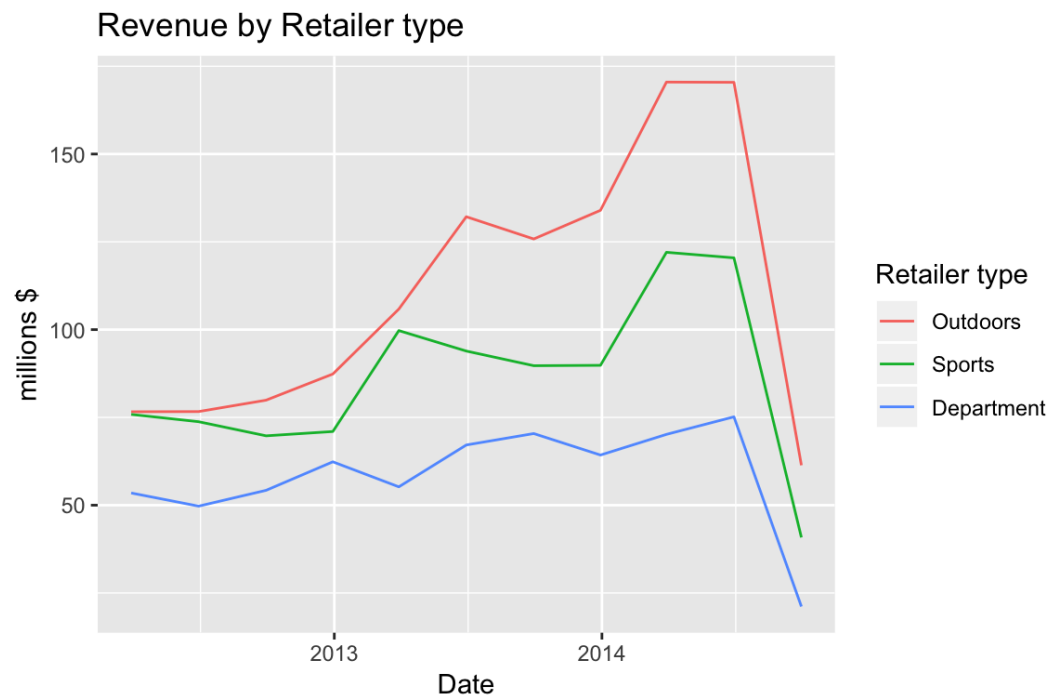
Consider one variable at a time

```
sales %>% group_by(Date, `Order method type`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = fct_reorder2(`Order method type`, Date, SumRev))) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year")  + ylab("millions $") + ggtitle("Revenue by Order method type") +
  labs(color = "Order method type")
```


Revenue by Order method type

Consider one variable at a time

```r
sales %>% group_by(Date, `Retailer type`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = fct_reorder2(`Retailer type`, Date, SumRev))) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year")  + ylab("millions $") + ggtitle("Revenue by
  Retailer type") +
  labs(color = "Retailer type")
```


Revenue by Retailer type

Add faceting (one dimension)

Reorder factor levels

```
sales <- sales %>%
  mutate(`Product line` = reorder(`Product line`, Revenue, function(x) sum(x)*-1),
         `Order method type` = reorder(`Order method type`, Revenue, function(x) sum(x)*-1),
         `Retailer type` = reorder(`Retailer type`, Revenue, function(x) sum(x)*-1))
```
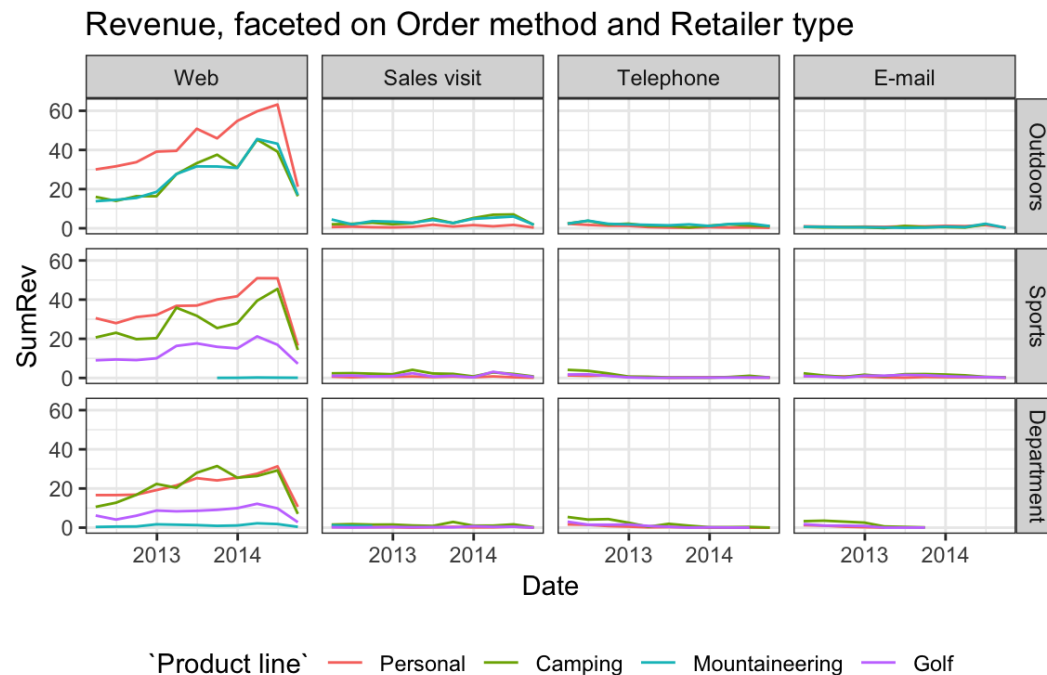
Add faceting (one dimension)

```r
sales %>% group_by(Date, `Product line`, `Order method type`) %>%
  summarize(SumRev = sum(Revenue))  %>%
  ggplot(aes(Date, SumRev, color = `Order method type`)) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  facet_wrap(~`Product line`) +  ylab("millions $") +
  theme(legend.position = "bottom") +
  ggtitle("Revenue, faceted on product line")
```
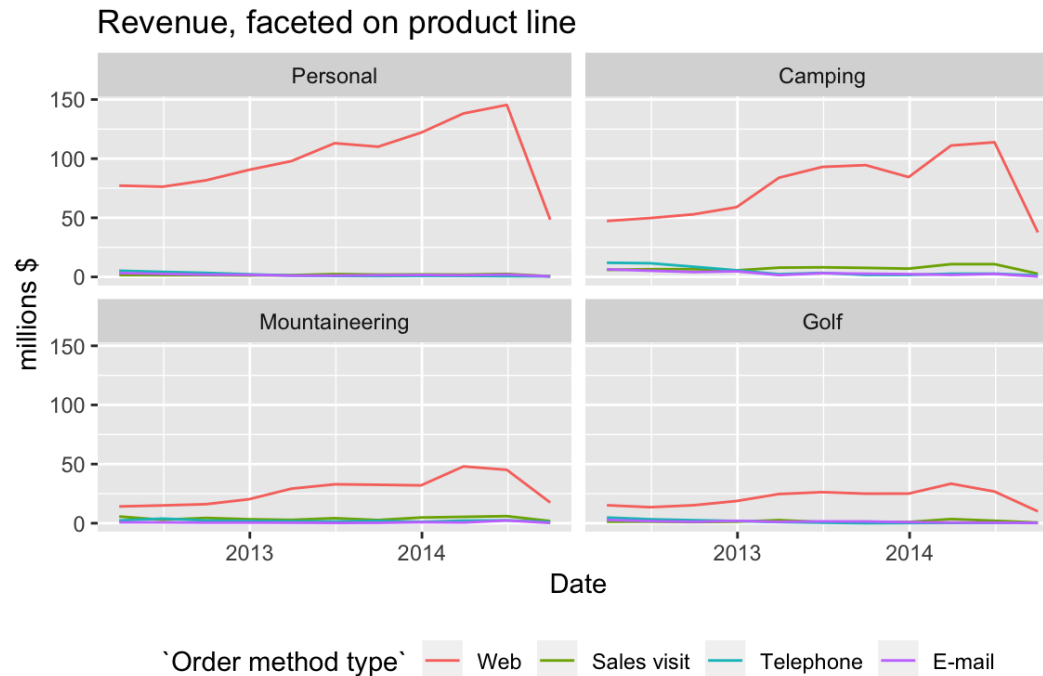
Faceting (two dimensions)

```r
sales %>% group_by(Date, `Product line`, `Retailer type`, `Order method type`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = `Product line`)) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  facet_grid(`Retailer type`~`Order method type`) +
  theme_bw() + theme(legend.position = "bottom") +
  ggtitle("Revenue, faceted on Order method and Retailer type")
```


Revenue, faceted on Order method and Retailer type

Experiment

```
sales %>% group_by(Date, `Product line`, `Retailer type`, `Order method type`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = `Retailer type`)) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  facet_grid(`Product line`~`Order method type`) +
  theme_bw() + theme(legend.position = "bottom") +
  ggtitle("Revenue, faceted on Order method and Product line")
```


Revenue, faceted on Order method and Product line

Experiment

```r
sales %>% group_by(Date, `Product line`, `Retailer type`, `Order method type`) %>%
  summarize(SumRev = sum(Revenue)) %>%
  ggplot(aes(Date, SumRev, color = `Order method type`)) + geom_line() +
  scale_x_date(date_labels = "%Y", date_breaks = "1 year") +
  facet_grid(`Product line`~`Retailer type`) +
  theme_bw() + theme(legend.position = "bottom") +
  ggtitle("Revenue, faceted on Retailer type and Product line")
```
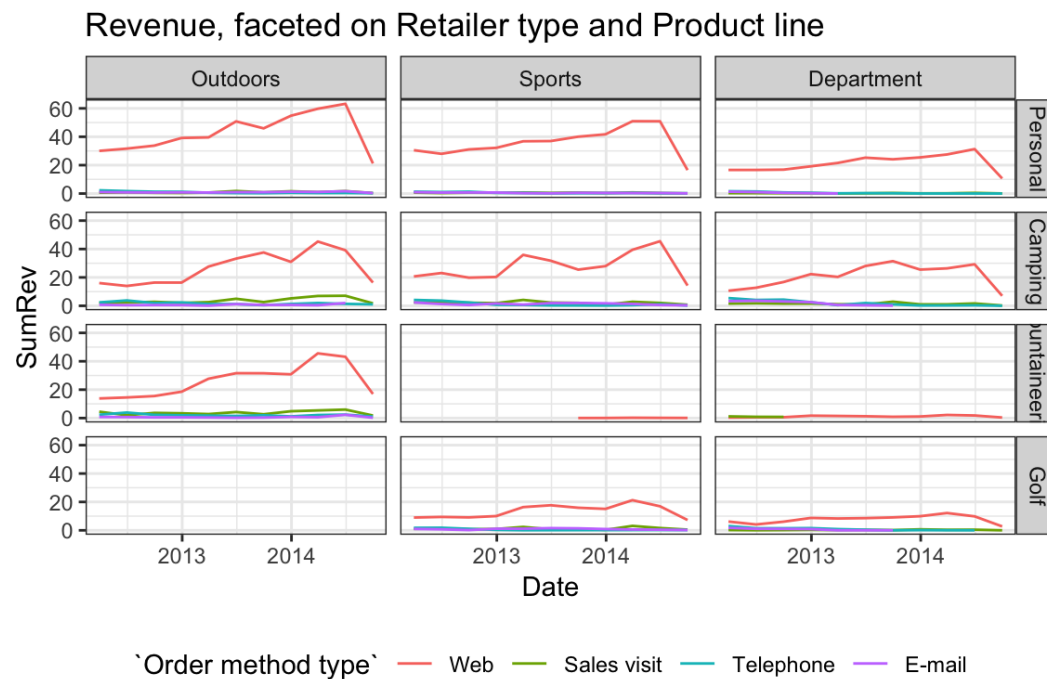
Scatterplot matrices

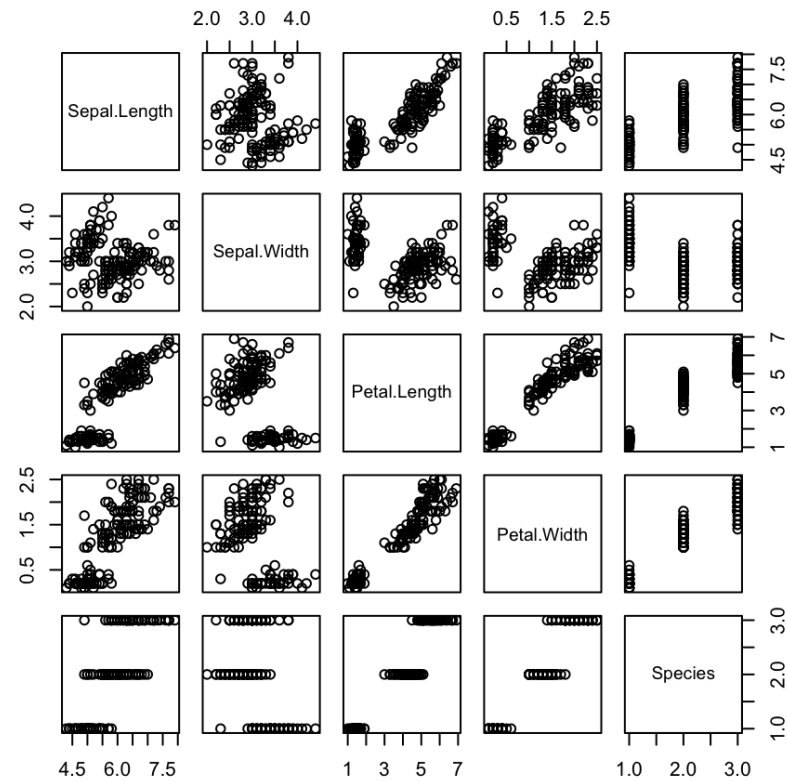Can't create with faceting in `ggplot2`

Options:

`plot()`
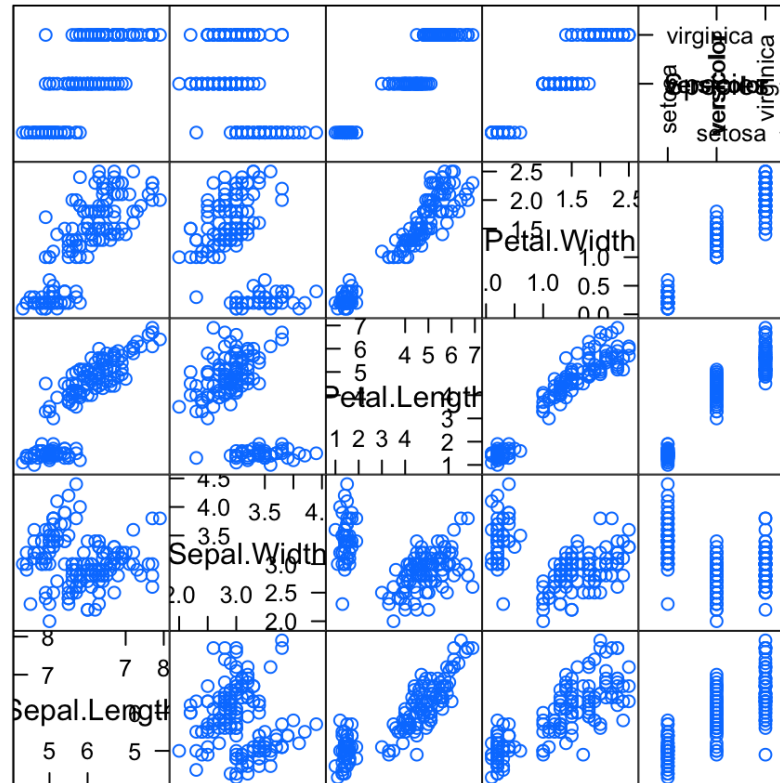
`lattice::splom()`

`GGally::ggpairs()`

Scatterplot matrices

```
plot(iris)
```
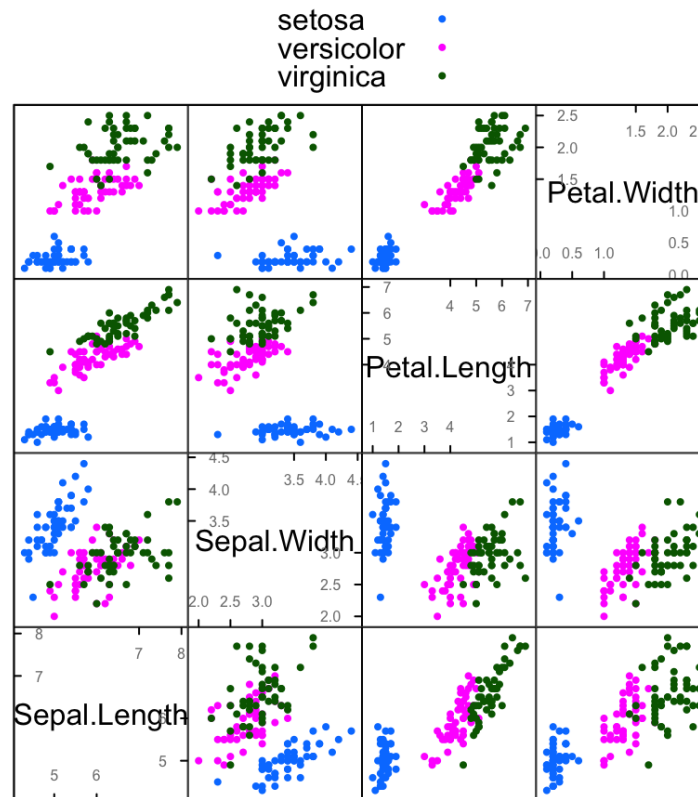
Scatterplot matrices

```
lattice::splom(iris)
```


Scatter Plot Matrix

Scatterplot matrices

```r
lattice::splom(iris[,1:4], group = iris$Species,
               par.settings = list(superpose.symbol = list(pch = 16, cex = .5)),
               axis.text.cex = .5, axis.text.col = "grey50",
               axis.line.tck = .5, auto.key = TRUE)
```


Scatter Plot Matrix

Scatterplot matrices

```
GGally::ggpairs(iris, mapping = ggplot2::aes(color = Species))
```