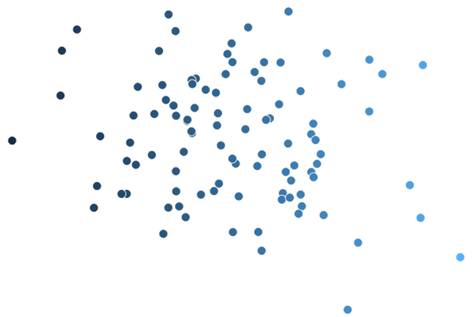
Grammar of Graphics

scales and coordinate systems

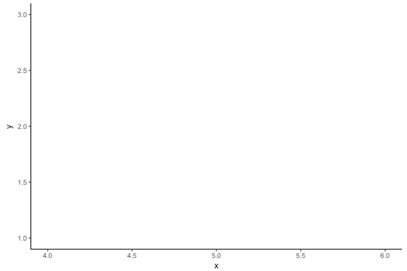
`slides/04_scales.pdf`

Building blocks

Layer(s)

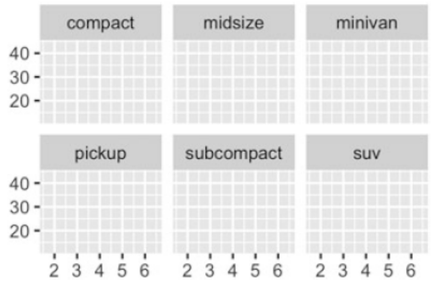


Scale(s)

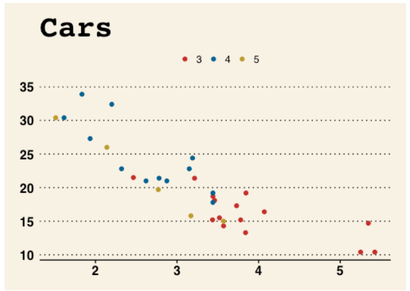


Coord

Facet

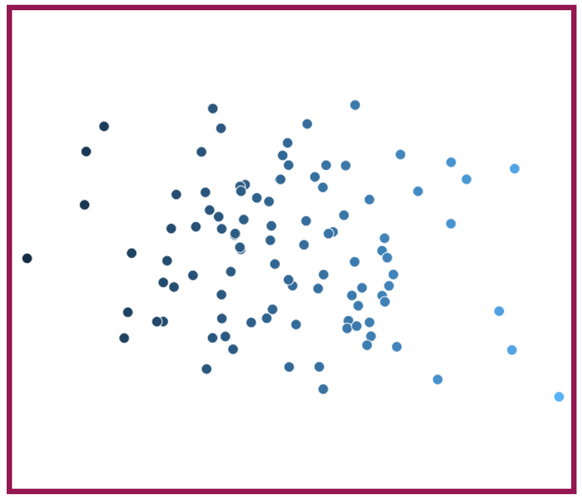


Theme



Layers

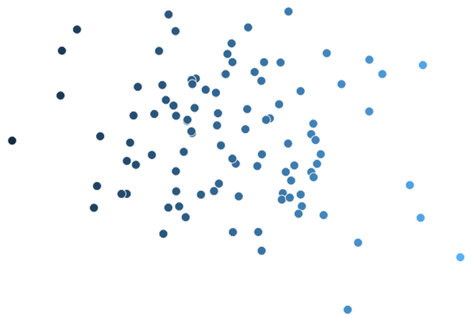
Each layer consists of:



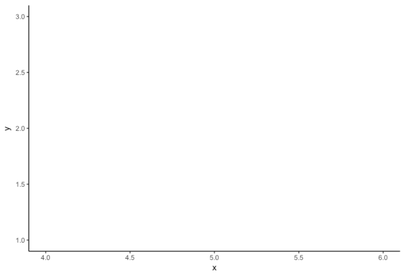
1. GEOM
2. AESTHETIC MAPPING
3. DATA
4. STAT
5. POSITION

Building blocks

Layer(s)

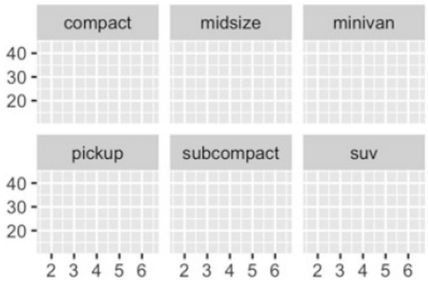


Scale(s)

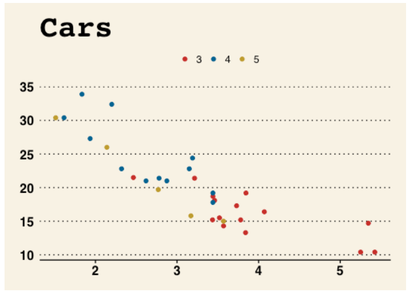


Coord

Facet



Theme

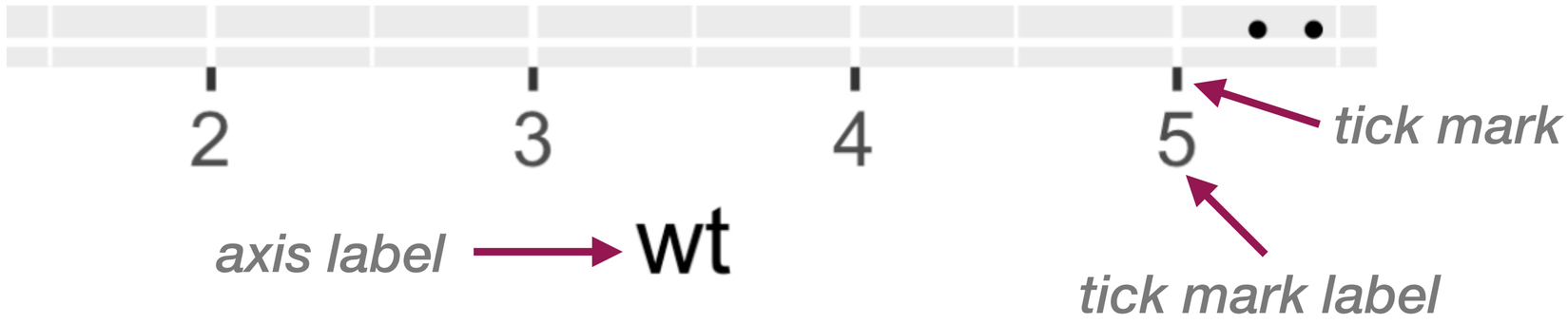


Scales

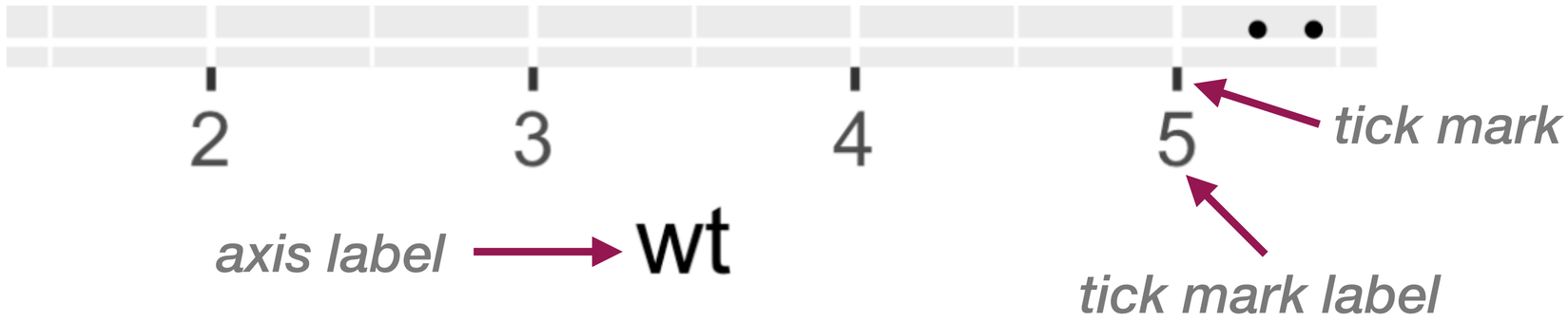
- One per aesthetic mapping
- The scale must match the data type (continuous or discrete)

MAPPING	SCALE
x →	`scale_x_date()`
y →	`scale_y_continuous()`
color →	`scale_color_manual()`
fill →	`scale_fill_viridis_c()`

x and y scales



Continuous x and y scales

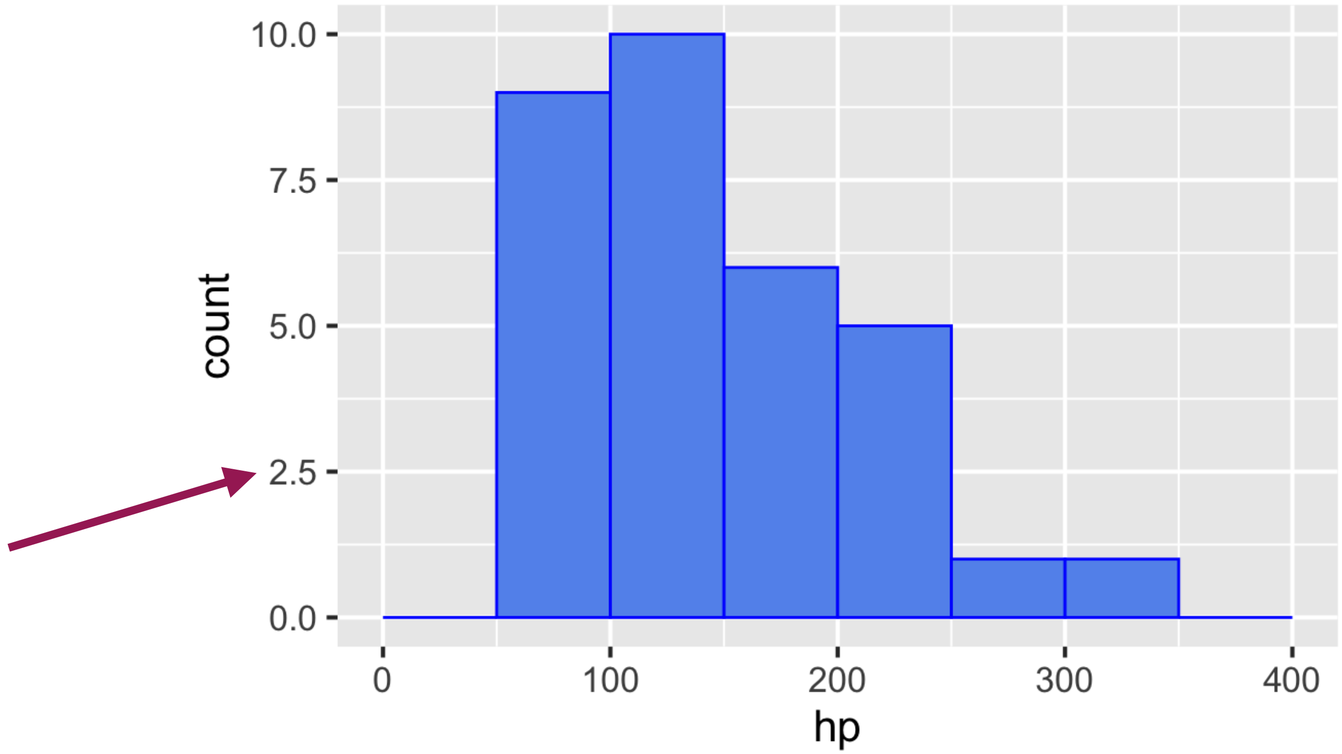


`+ scale_x_continuous()` or `+ scale_y_continuous()`

- change axis label with `name =` or `labs(x = ...)`
- set range of axis with `limits =`
- choose tick mark locations with `breaks =`
- format tick mark labels `labels =` (rare, often a function)

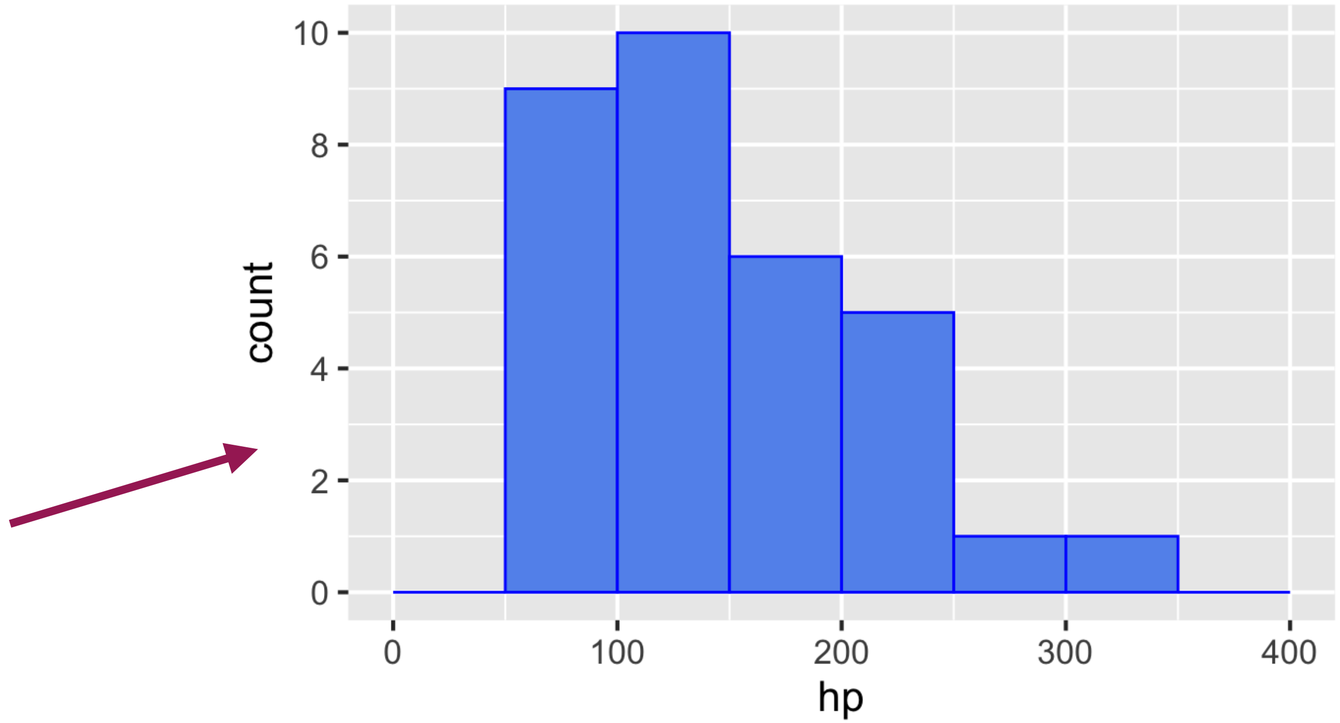
Problematic y-axis

```
ggplot(mtcars, aes(x = hp)) +
  geom_histogram(breaks = seq(0, 400, 50),
                 color = "blue", fill = "cornflowerblue")
```



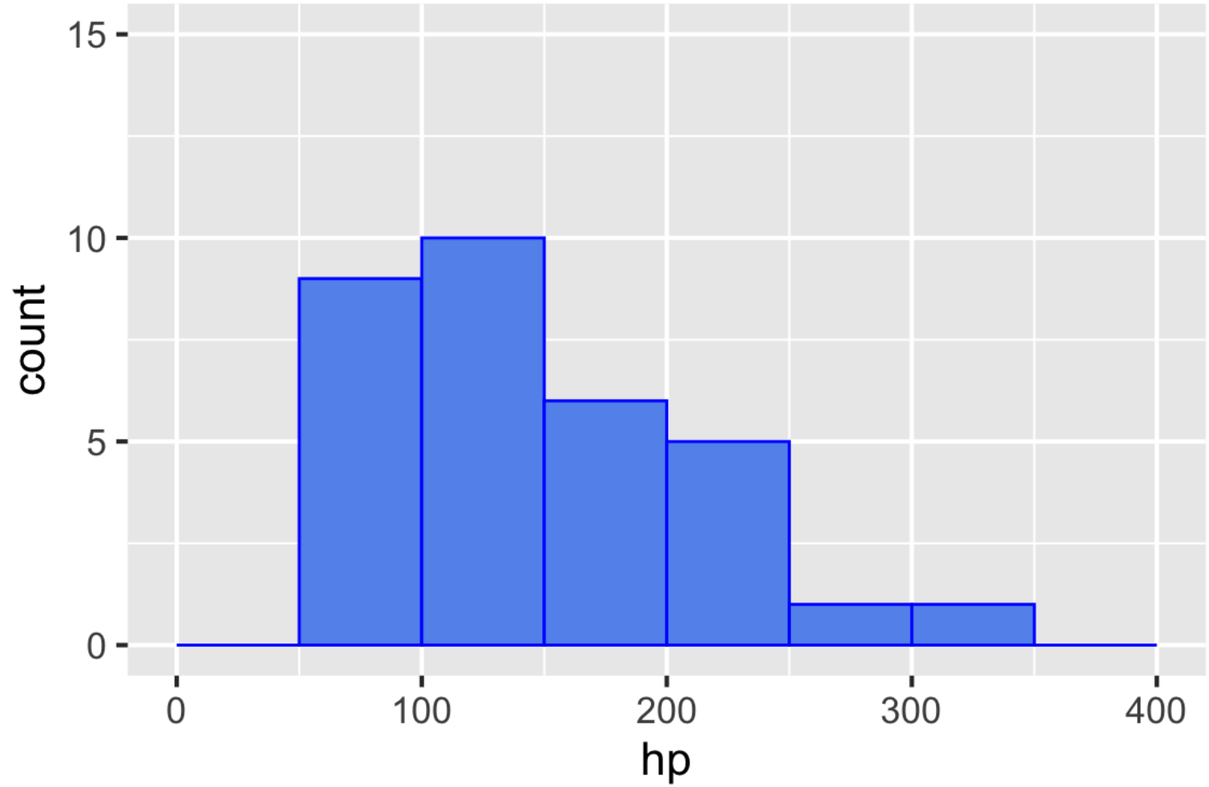
Change scale breaks

```
ggplot(mtcars, aes(x = hp)) +
  geom_histogram(breaks = seq(0, 400, 50),
                 color = "blue", fill = "cornflowerblue") +
  scale_y_continuous(breaks = seq(0, 10, 2))
```



Change scale breaks

```
ggplot(mtcars, aes(x = hp)) +
  geom_histogram(breaks = seq(0, 400, 50),
                 color = "blue", fill = "cornflowerblue") +
  scale_y_continuous(limits = c(0, 15))
```



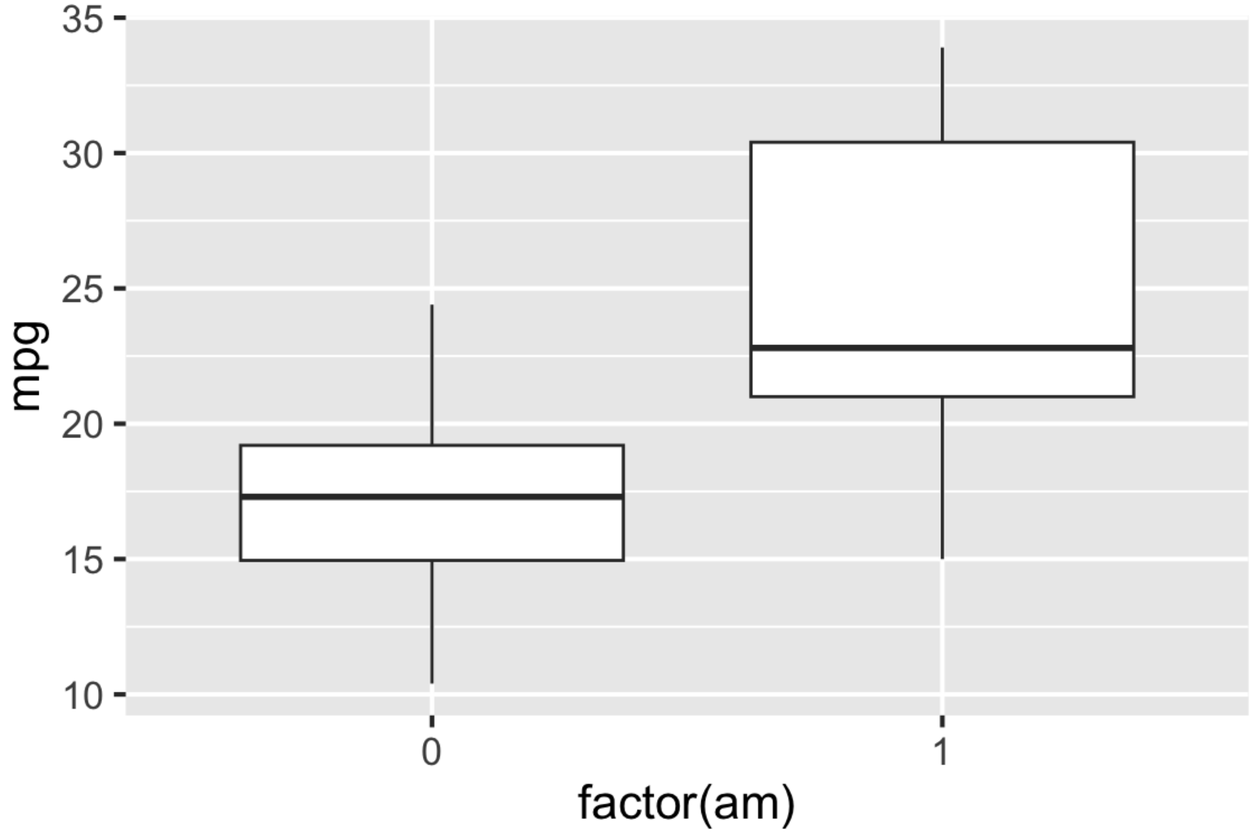
Discrete x and y scales

`+ scale_x_discrete()` or `+ scale_y_discrete()`

- change axis label with `name =` or `labs(x = ...)`
- change tick mark labels `labels =`

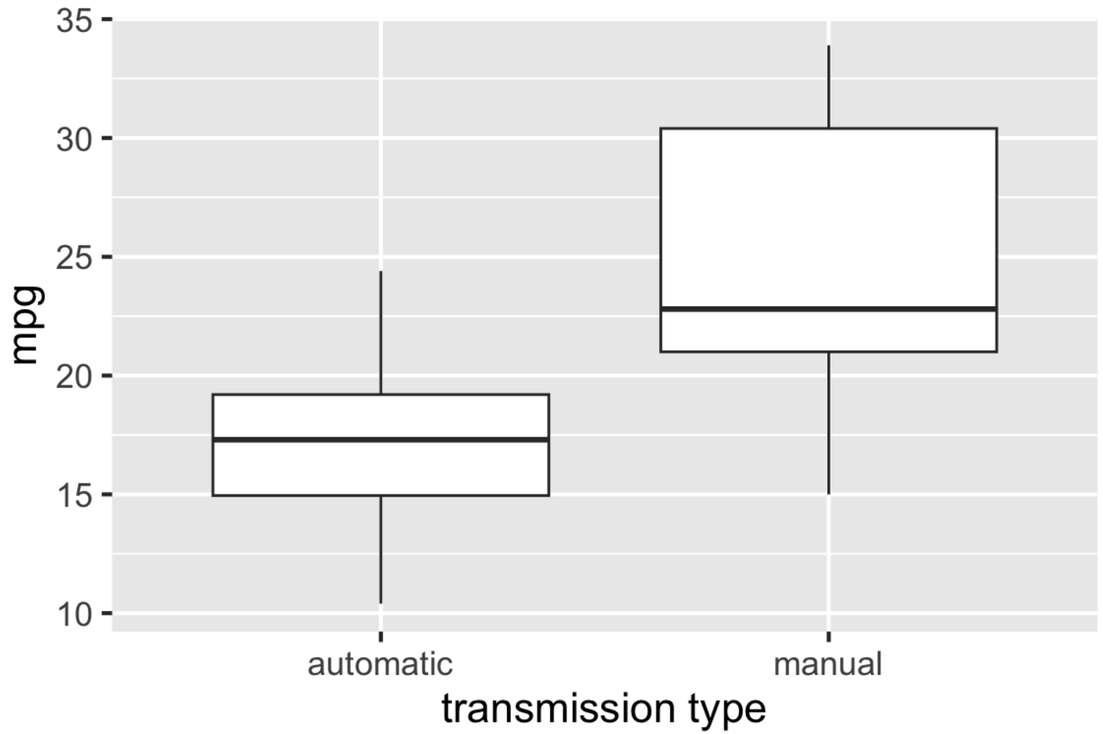
Unclear tick mark labels

```
ggplot(mtcars, aes(x = factor(am), y = mpg)) +
  geom_boxplot()
```



Rename tick mark labels

```
ggplot(mtcars, aes(x = factor(am), y = mpg)) +
  geom_boxplot() +
  scale_x_discrete(name = "transmission type",
                   labels = c("0" = "automatic", "1" = "manual"))
```



Discrete vs. continuous

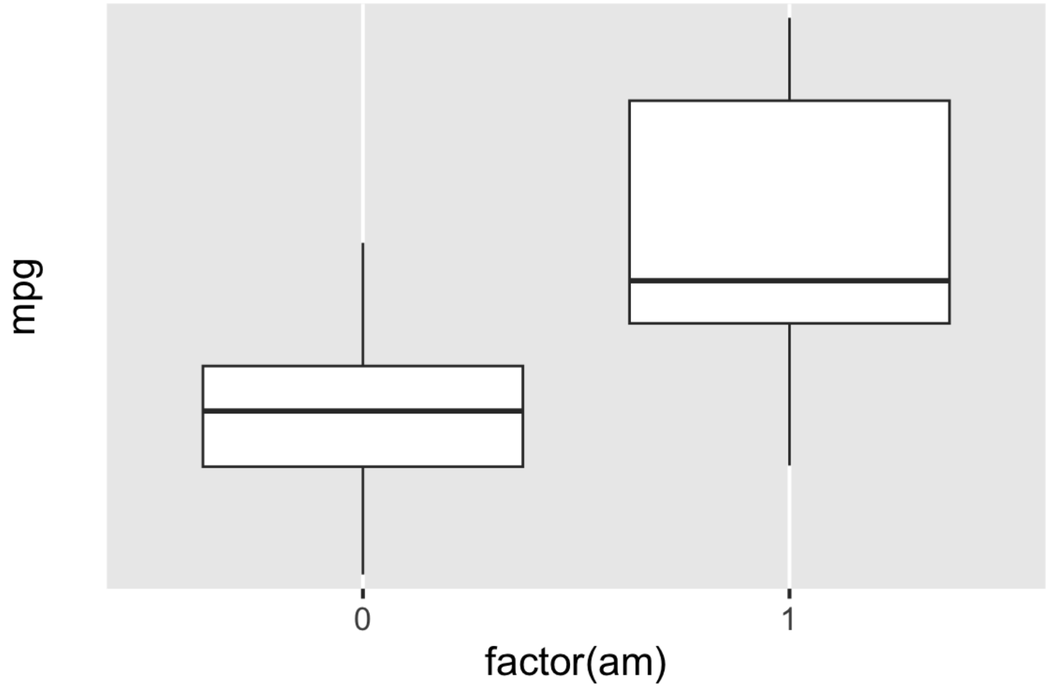

```
ggplot(mtcars, aes(x = factor(am), y = mpg)) +
  geom_boxplot() +
  scale_x_continuous(name = "transmission type",
                     labels = c("0" = "automatic", "1" = "manual"))
```

```
Error in `train_continuous()`:
! Discrete value supplied to a continuous scale
```

Discrete vs. continuous


```
ggplot(mtcars, aes(x = factor(am), y = mpg)) +
  geom_boxplot() +
  scale_y_discrete(limits = c(0, 40))
```

```
Warning: Continuous limits supplied to discrete scale.
ℹ Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```

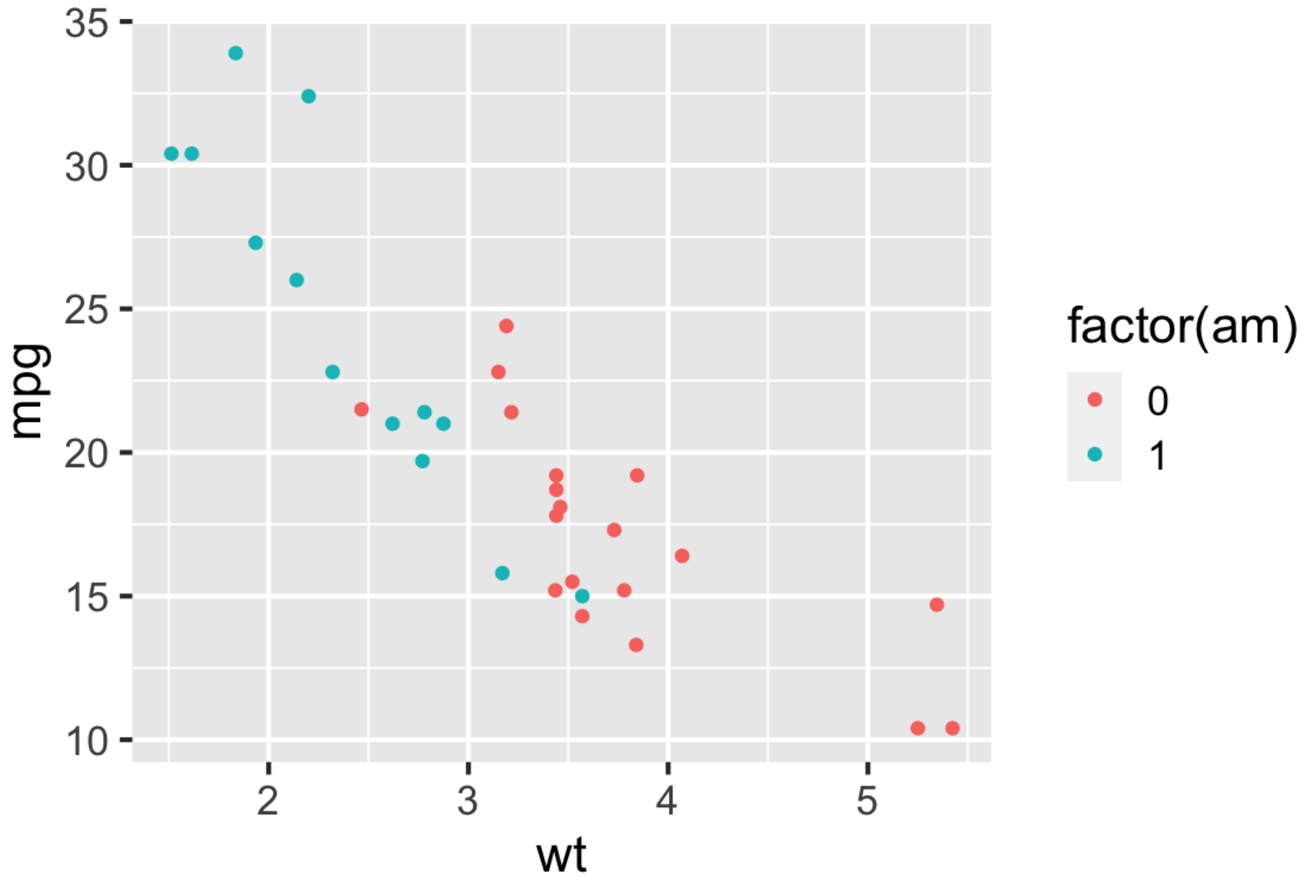


Color / fill scales

- continuous data --> continuous color / fill scale
- discrete data --> discrete color / fill scale
- color / fill scales only apply to aesthetic mappings
- for "constant" colors, use `color =` or `fill =`

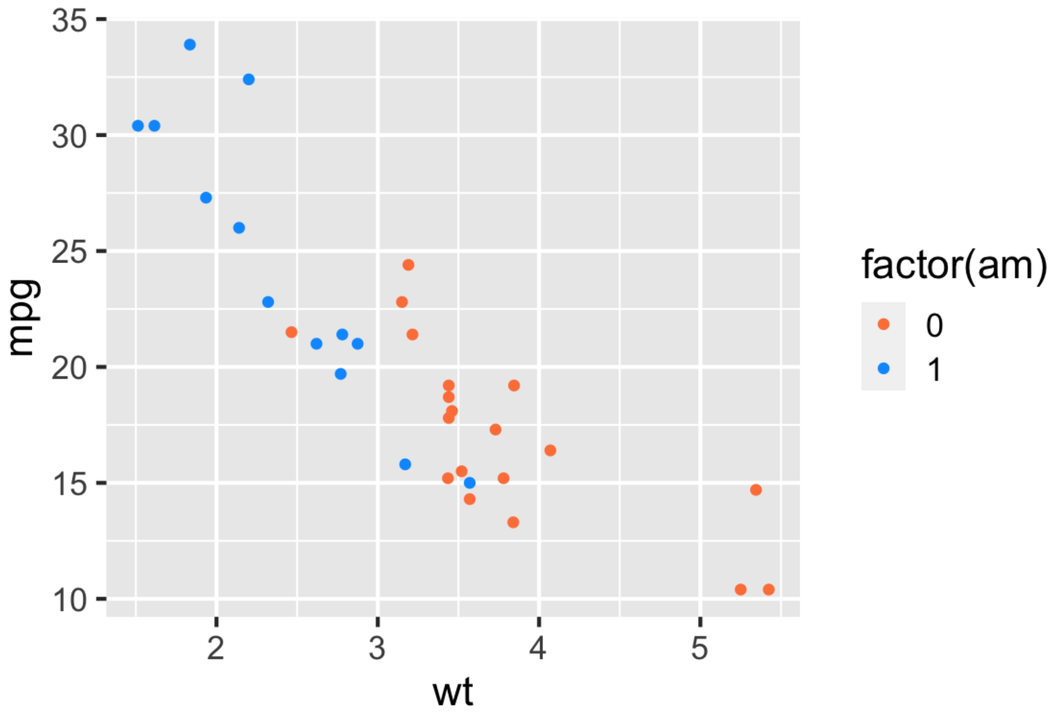
Default discrete colors

```
ggplot(mtcars, aes(x = wt, y = mpg, color = factor(am))) +
  geom_point()
```



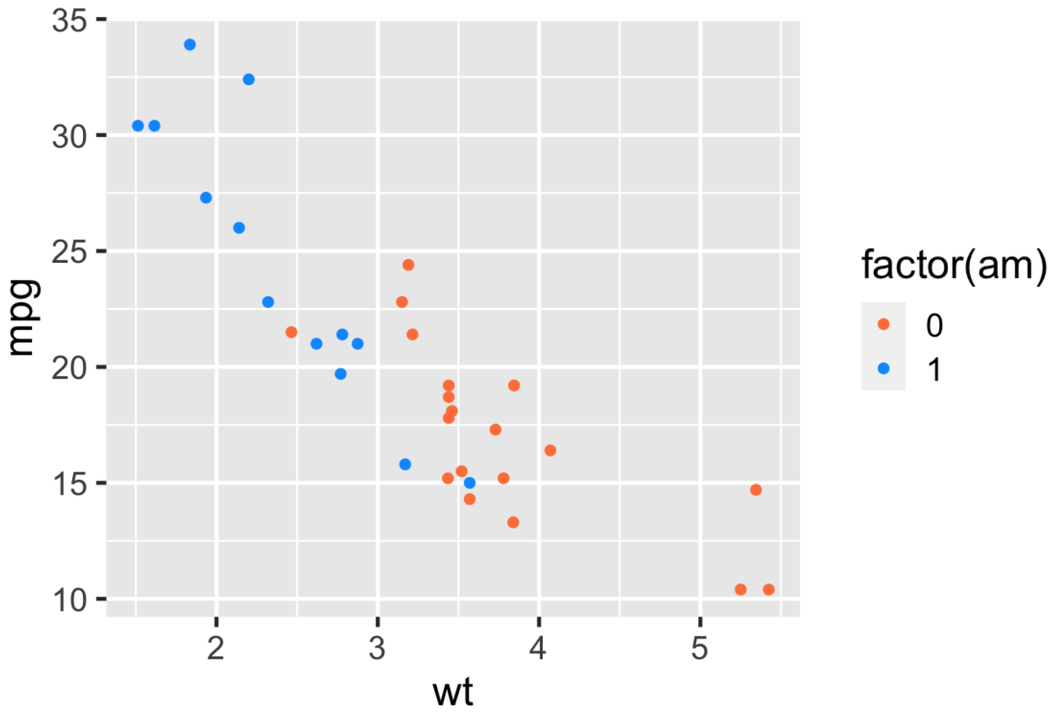
Change discrete colors (manual)

```
ggplot(mtcars, aes(x = wt, y = mpg, color = factor(am))) +
  geom_point() +
  scale_color_manual(values = c("#FF8146", "#009BFF"))
```



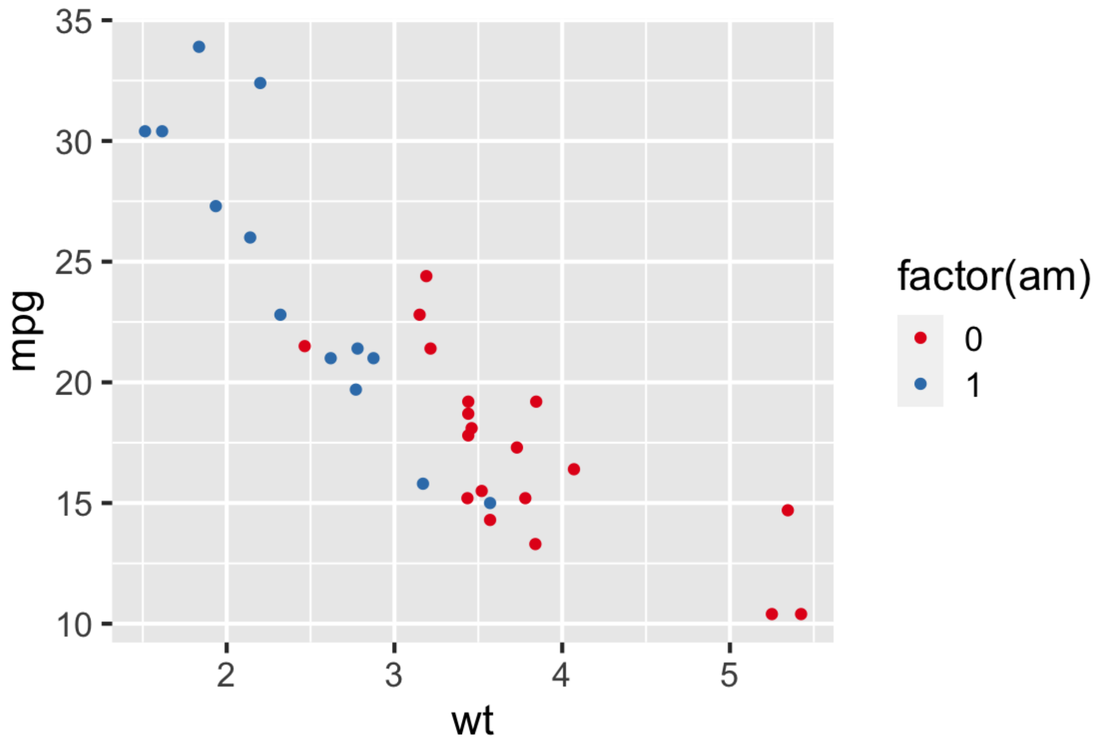
RStudio view

```r
ggplot(mtcars, aes(x = wt, y = mpg, color = factor(am))) +
  geom_point() +
  scale_color_manual(values = c("#FF8146", "#009BFF"))
```



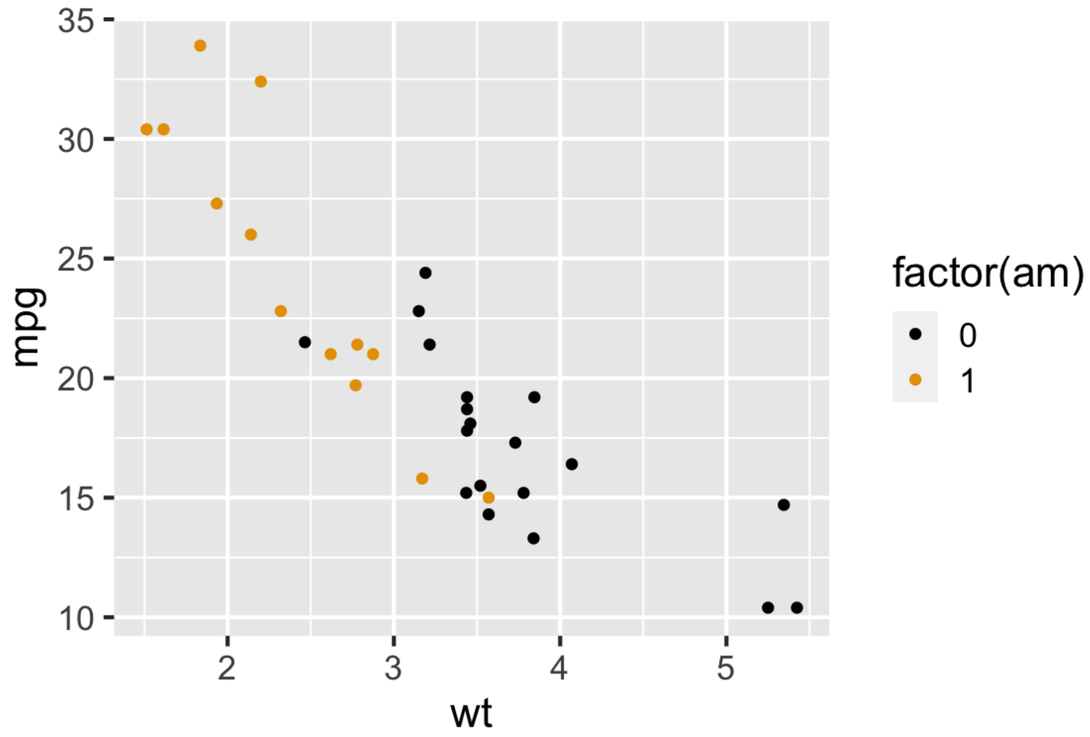
Change discrete colors (prebuilt)

```
ggplot(mtcars, aes(x = wt, y = mpg, color = factor(am))) +
  geom_point() +
  scale_color_brewer(palette = "Set1")
```



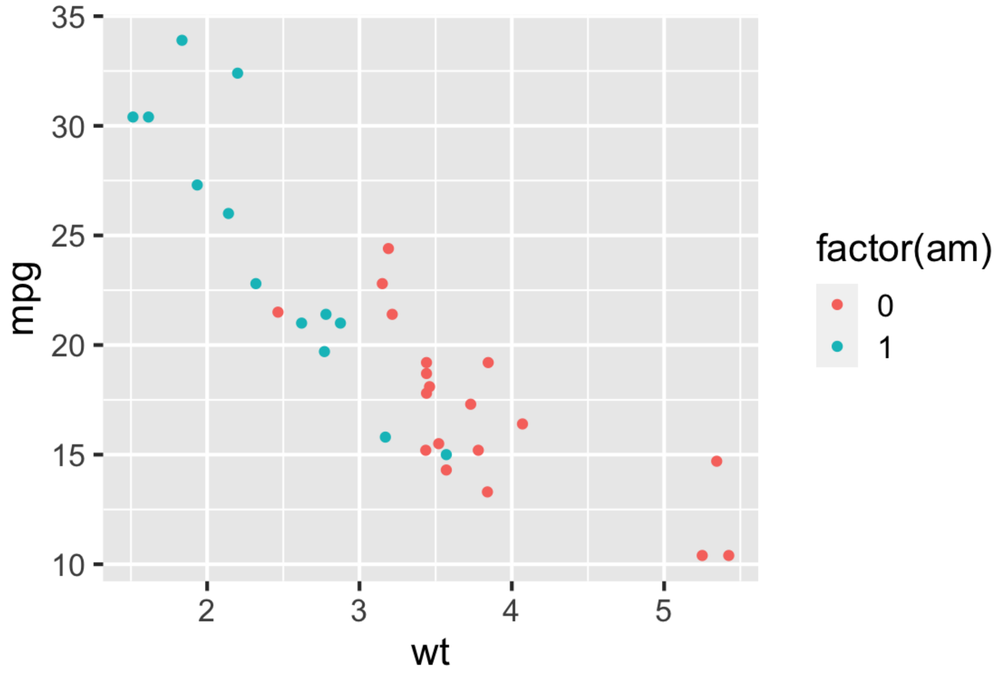
Change discrete colors (prebuilt)

```
library(ggthemes)
ggplot(mtcars, aes(x = wt, y = mpg, color = factor(am))) +
  geom_point() +
  scale_color_colorblind()
```



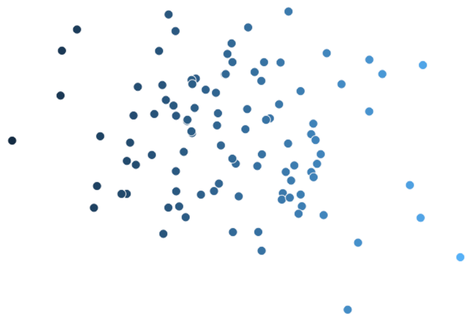
Mixing up color and fill


```
ggplot(mtcars, aes(x = wt, y = mpg,
                   color = factor(am))) +
  geom_point() +
  scale_fill_manual(values = c("orange", "black"))
```

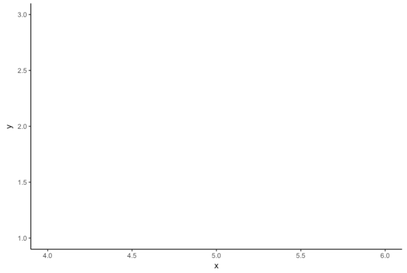


Building blocks

Layer(s)

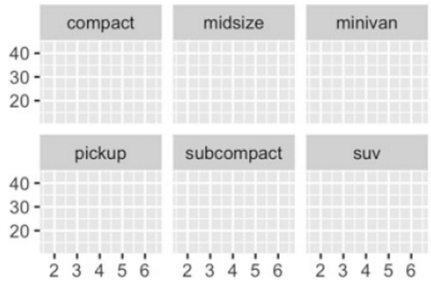


Scale(s)

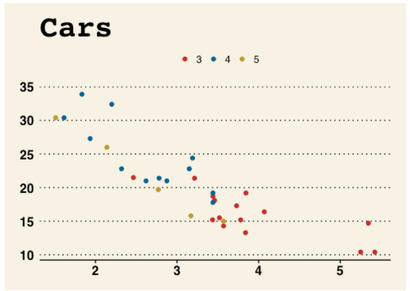


Coord

Facet



Theme

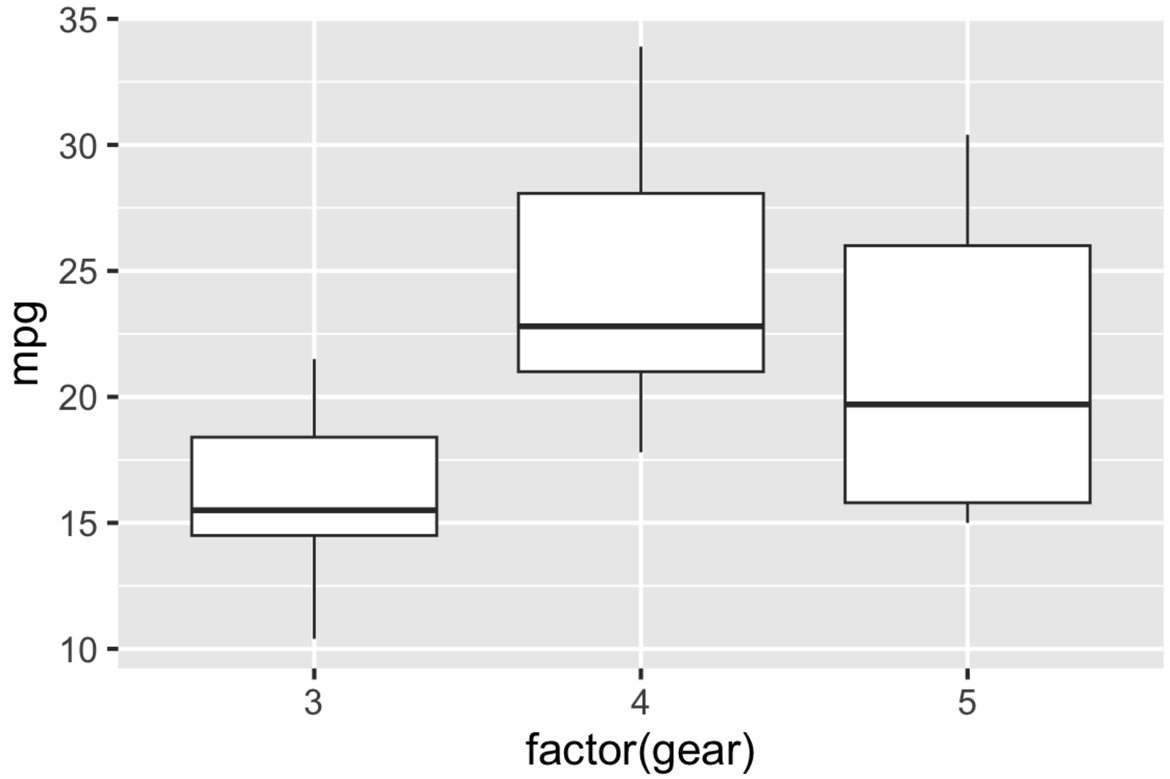


Coordinate systems

- The default coordinate system is `coord_cartesian()`.
- It is rare to need to change it.
- `coord_flip()` is a common option that keeps the Cartesian system but switches x and y. Note that the x and y scales do not change, which is confusing.
- It is preferable to switch the x and y aesthetics if possible rather than use `coord_flip()`.

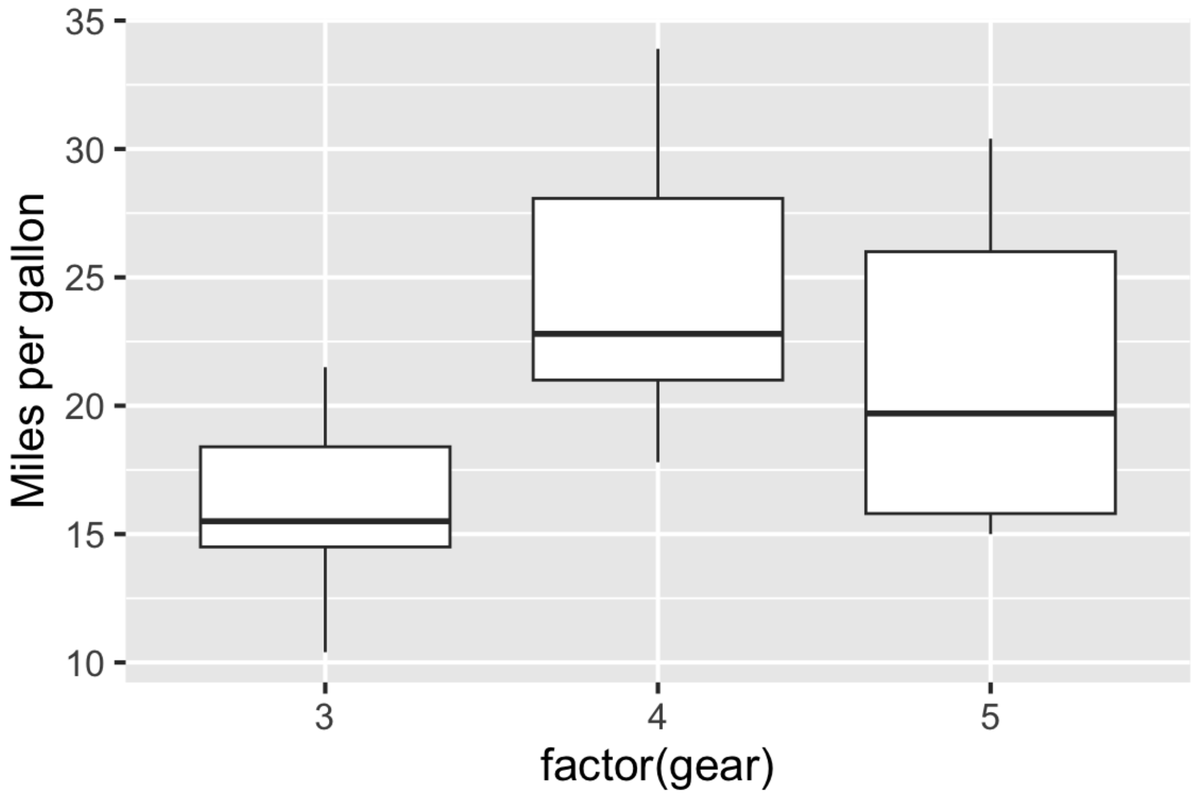
Flipping x and y axes

```
ggplot(mtcars, aes(x = mpg, y = factor(gear))) +
  geom_boxplot() +
  coord_flip()
```



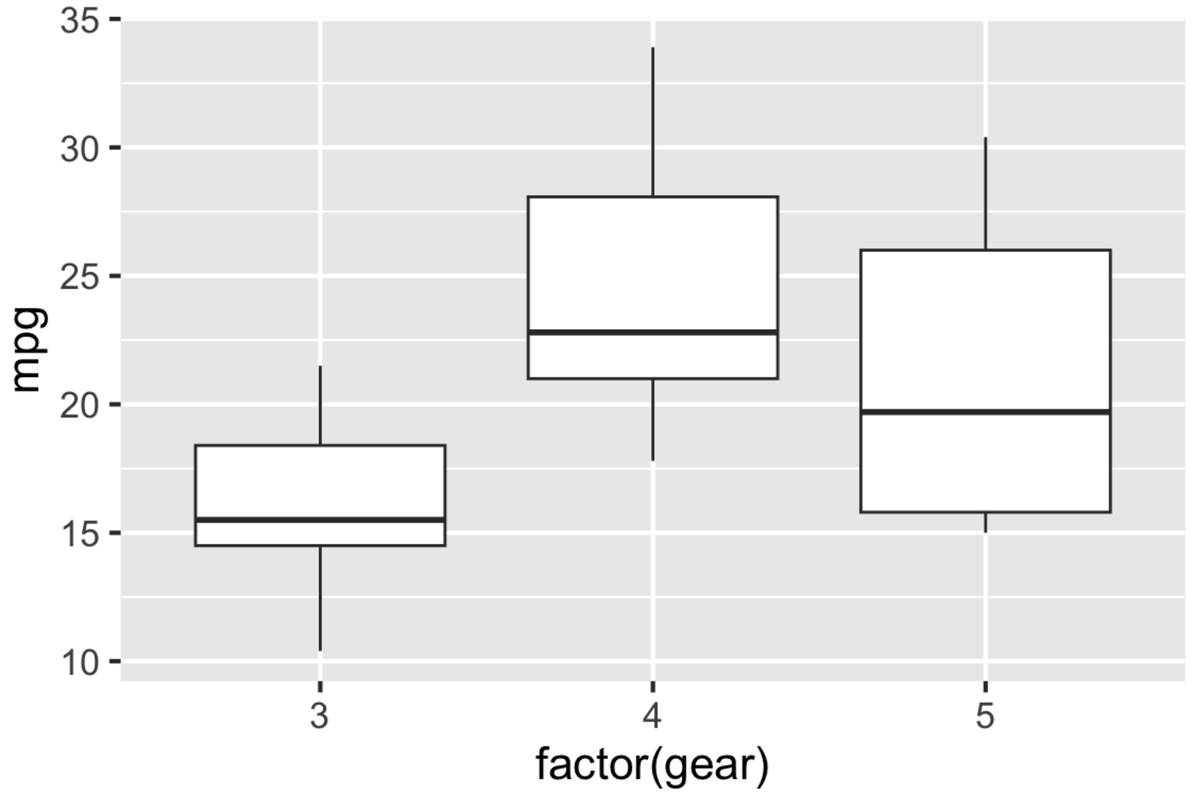
... but scales stay with original mappings

```
ggplot(mtcars, aes(x = mpg, y = factor(gear))) + geom_boxplot() +
  scale_x_continuous(name = "Miles per gallon") +
  coord_flip()
```



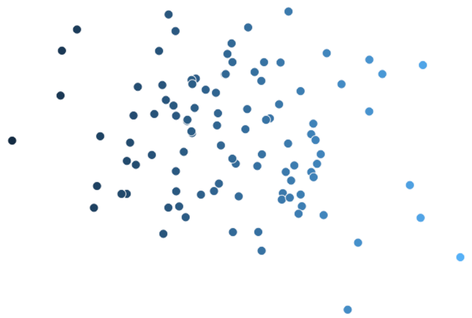
A better approach

```
ggplot(mtcars, aes(x = factor(gear), y = mpg)) +
  geom_boxplot()
```

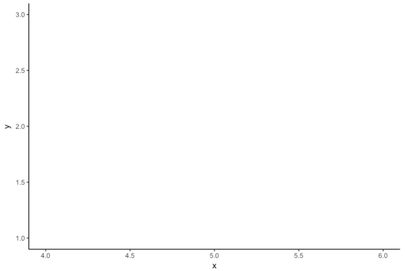


Building blocks

Layer(s)

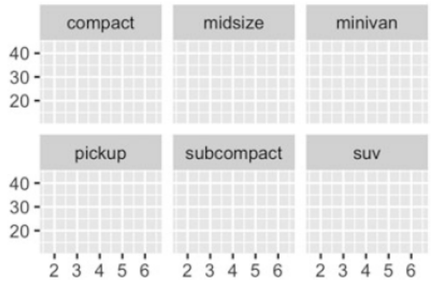


Scale(s)

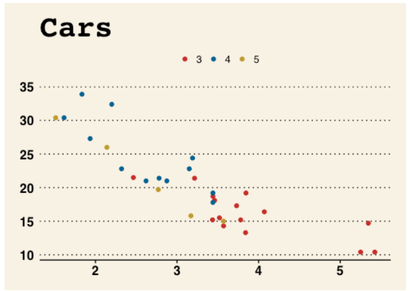


Coord

Facet



Theme



Themes

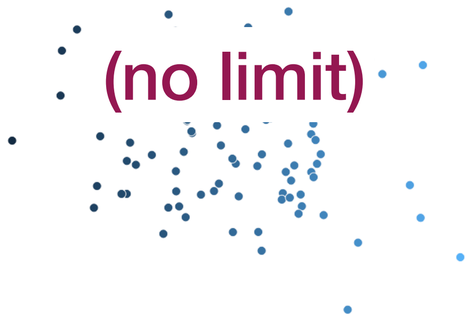
- The default theme is `theme_grey(base_size = 11)`
- To increase the font size of all text elements, increate the base font size: `+ theme_grey(14)`
- Other common built-in themes:

  `theme_bw()`
  `theme_classic()`
  `theme_linedraw()`
  `theme_void()`

- There are lots of other themes in the **`ggthemes`** package

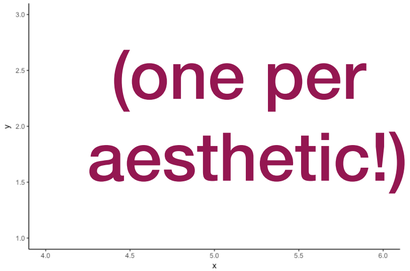
How many?

Layer(s)



(no limit)

Scale(s)

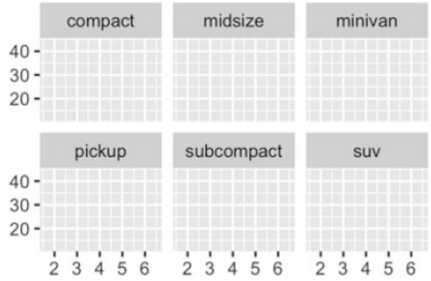


(one per aesthetic!)

Coord

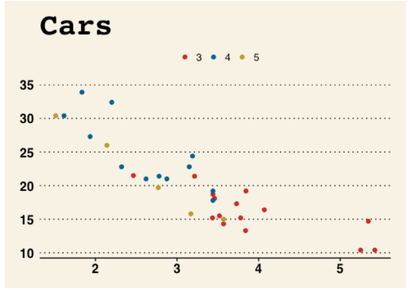
(one!)

Facet



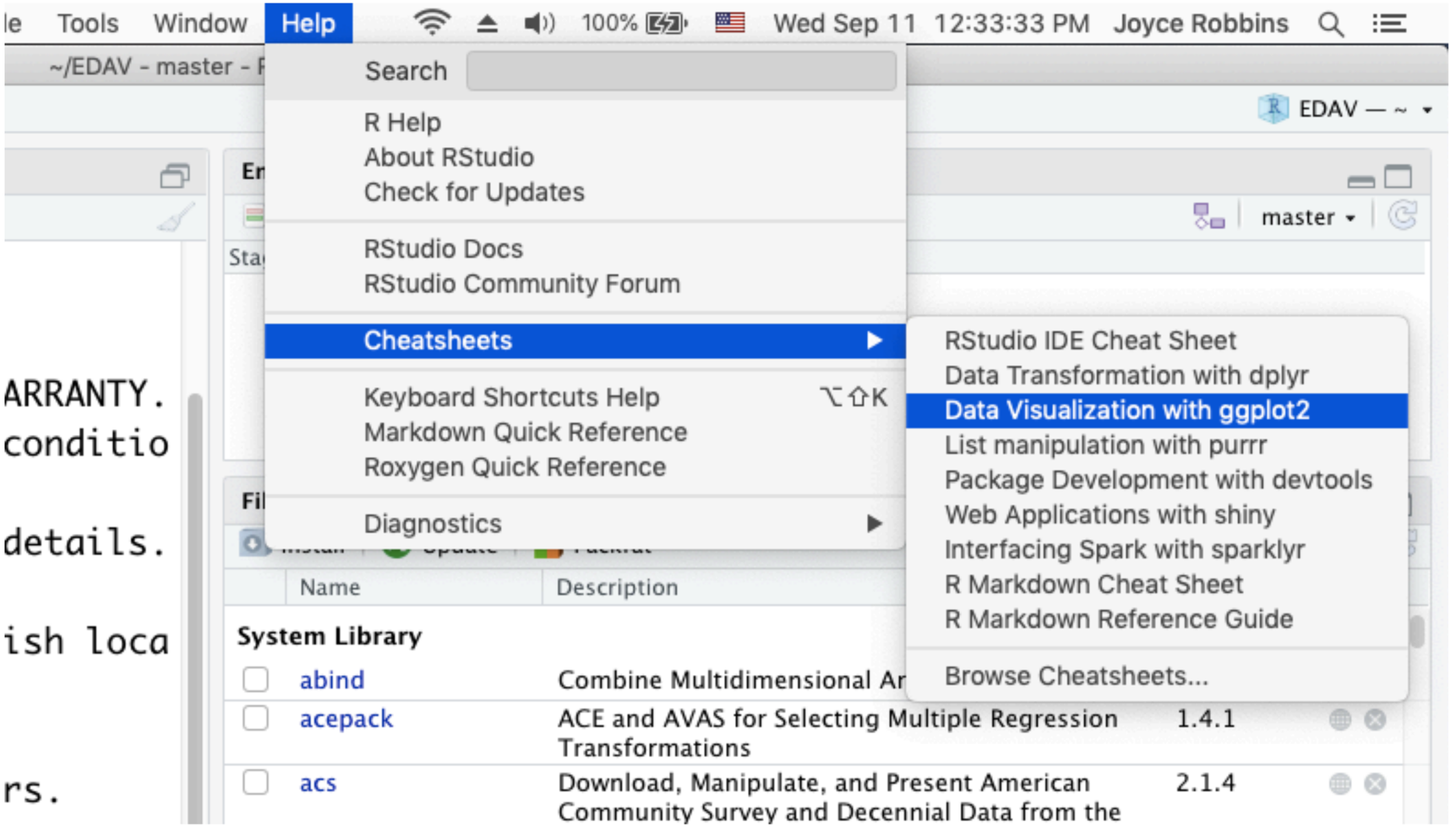
(one!)

Theme



(one, sort of)

Cheatsheet


Tools
Window
Help
100%
Wed Sep 11 12:33:33 PM
Joyce Robbins
~/EDAV - master - R
EDAV — ~
master
Search
R Help
About RStudio
Check for Updates
RStudio Docs
RStudio Community Forum
Cheatsheets
Keyboard Shortcuts Help
Markdown Quick Reference
Roxygen Quick Reference
Diagnostics
RStudio IDE Cheat Sheet
Data Transformation with dplyr
Data Visualization with ggplot2
List manipulation with purrr
Package Development with devtools
Web Applications with shiny
Interfacing Spark with sparklyr
R Markdown Cheat Sheet
R Markdown Reference Guide
Browse Cheatsheets...
ARRANTY.
conditio
details.
ish loca
rs.
Name
Description
System Library
abind
Combine Multidimensional Ar
acepack
ACE and AVAS for Selecting Multiple Regression Transformations
1.4.1
acs
Download, Manipulate, and Present American Community Survey and Decennial Data from the
2.1.4

Data Visualization with ggplot2 : : CHEAT SHEET

Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data** set, a **coordinate system**, and geoms—visual marks that represent data points.

data **geom** x = F, y = A **coordinate system** **plot**

To display values, map variables in the data to visual properties of the geom (**aesthetics**) like **size**, **color**, and **x** and **y** locations.

data **geom** x = F, y = A, color = F, size = A **coordinate system** **plot**

Complete the template below to build a graph.

```
ggplot (data = <DATA>) +                                  required
<GEOM_FUNCTION> (mapping = aes(<MAPPINGS>),
 stat = <STAT>, position = <POSITION> ) +                 Not required, sensible defaults supplied
<COORDINATE_FUNCTION> +
<FACET_FUNCTION> +
<SCALE_FUNCTION> +
<THEME_FUNCTION>
```

ggplot(data = mpg, **aes(**x = cty, y = hwy**))** Begins a plot that you finish by adding layers to. Add one geom function per layer.

aesthetic mappings · data · geom

qplot(x = cty, y = hwy, data = mpg, geom = "point"**)** Creates a complete plot with given data, geom, and mappings. Supplies many useful defaults.

last_plot() Returns the last plot

ggsave("plot.png", width = 5, height = 5) Saves last plot as 5' x 5' file named "plot.png" in working directory. Matches file type to file extension.

Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

GRAPHICAL PRIMITIVES

```
a <- ggplot(economics, aes(date, unemploy))
b <- ggplot(seals, aes(x = long, y = lat))
```

a + geom_blank()
(Useful for expanding limits)

b + geom_curve(aes(yend = lat + 1, xend=long+1,curvature=z)**)** - x, xend, y, yend, alpha, angle, color, curvature, linetype, size

a + geom_path(lineend="butt", linejoin="round", linemitre=1**)**
x, y, alpha, color, group, linetype, size

a + geom_polygon(aes(group = group)**)**
x, y, alpha, color, fill, group, linetype, size

b + geom_rect(aes(xmin = long, ymin=lat, xmax= long + 1, ymax = lat + 1)**)** - xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size

a + geom_ribbon(aes(ymin=unemploy - 900, ymax=unemploy + 900)**)** - x, ymax, ymin, alpha, color, fill, group, linetype, size

LINE SEGMENTS

common aesthetics: x, y, alpha, color, linetype, size

b + geom_abline(aes(intercept=0, slope=1)**)**
b + geom_hline(aes(yintercept = lat)**)**
b + geom_vline(aes(xintercept = long)**)**

b + geom_segment(aes(yend=lat+1, xend=long+1)**)**
b + geom_spoke(aes(angle = 1:1155, radius = 1)**)**

ONE VARIABLE continuous

```
c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)
```

c + geom_area(stat = "bin")
x, y, alpha, color, fill, linetype, size

c + geom_density(kernel = "gaussian"**)**
x, y, alpha, color, fill, group, linetype, size, weight

c + geom_dotplot()
x, y, alpha, color, fill

c + geom_freqpoly() x, y, alpha, color, group, linetype, size

c + geom_histogram(binwidth = 5**)** x, y, alpha, color, fill, linetype, size, weight

c2 + geom_qq(aes(sample = hwy)**)** x, y, alpha, color, fill, linetype, size, weight

discrete

```
d <- ggplot(mpg, aes(fl))
```

d + geom_bar()
x, alpha, color, fill, linetype, size, weight

TWO VARIABLES

continuous x , continuous y

```
e <- ggplot(mpg, aes(cty, hwy))
```

e + geom_label(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)** x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

e + geom_jitter(height = 2, width = 2**)**
x, y, alpha, color, fill, shape, size

e + geom_point(), x, y, alpha, color, fill, shape, size, stroke

e + geom_quantile(), x, y, alpha, color, group, linetype, size, weight

e + geom_rug(sides = "bl"**)**, x, y, alpha, color, linetype, size

e + geom_smooth(method = lm**)**, x, y, alpha, color, fill, group, linetype, size, weight

e + geom_text(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)**, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

discrete x , continuous y

```
f <- ggplot(mpg, aes(class, hwy))
```

f + geom_col(), x, y, alpha, color, fill, group, linetype, size

f + geom_boxplot(), x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight

f + geom_dotplot(binaxis = "y", stackdir = "center"**)**, x, y, alpha, color, fill, group

f + geom_violin(scale = "area"**)**, x, y, alpha, color, fill, group, linetype, size, weight

discrete x , discrete y

```
g <- ggplot(diamonds, aes(cut, color))
```

g + geom_count(), x, y, alpha, color, fill, shape, size, stroke

continuous bivariate distribution

```
h <- ggplot(diamonds, aes(carat, price))
```

h + geom_bin2d(binwidth = c(0.25, 500)**)**
x, y, alpha, color, fill, linetype, size, weight

h + geom_density2d()
x, y, alpha, colour, group, linetype, size

h + geom_hex()
x, y, alpha, colour, fill, size

continuous function

```
i <- ggplot(economics, aes(date, unemploy))
```

i + geom_area()
x, y, alpha, color, fill, linetype, size

i + geom_line()
x, y, alpha, color, group, linetype, size

i + geom_step(direction = "hv")
x, y, alpha, color, group, linetype, size

visualizing error

```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))
```

j + geom_crossbar(fatten = 2**)**
x, y, ymax, ymin, alpha, color, fill, group, linetype, size

j + geom_errorbar(), x, ymax, ymin, alpha, color, group, linetype, size, width (also **geom_errorbarh()**)

j + geom_linerange()
x, ymin, ymax, alpha, color, group, linetype, size

j + geom_pointrange()
x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, size

maps

```
data <- data.frame(murder = USArrests$Murder,
state = tolower(rownames(USArrests)))
map <- map_data("state")
k <- ggplot(data, aes(fill = murder))
```

k + geom_map(aes(map_id = state), map = map**)**
+ expand_limits(x = map$long, y = map$lat**)**,
map_id, alpha, color, fill, linetype, size

THREE VARIABLES

```
seals$z <- with(seals, sqrt(delta_long^2 + delta_lat^2)); l <- ggplot(seals, aes(long, lat))
```

l + geom_contour(aes(z = z))
x, y, z, alpha, colour, group, linetype, size, weight

l + geom_raster(aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE**)**
x, y, alpha, fill

l + geom_tile(aes(fill = z)**)**, x, y, alpha, color, fill, linetype, size, width

Code style

Complete the template below to build a graph.

```
ggplot (data = <DATA>) +
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
 stat =<STAT>, position = <POSITION> ) +
<COORDINATE_FUNCTION>+
<FACET_FUNCTION>+
<SCALE_FUNCTION>+
<THEME_FUNCTION>
```

required

Not required, sensible defaults supplied

Code style

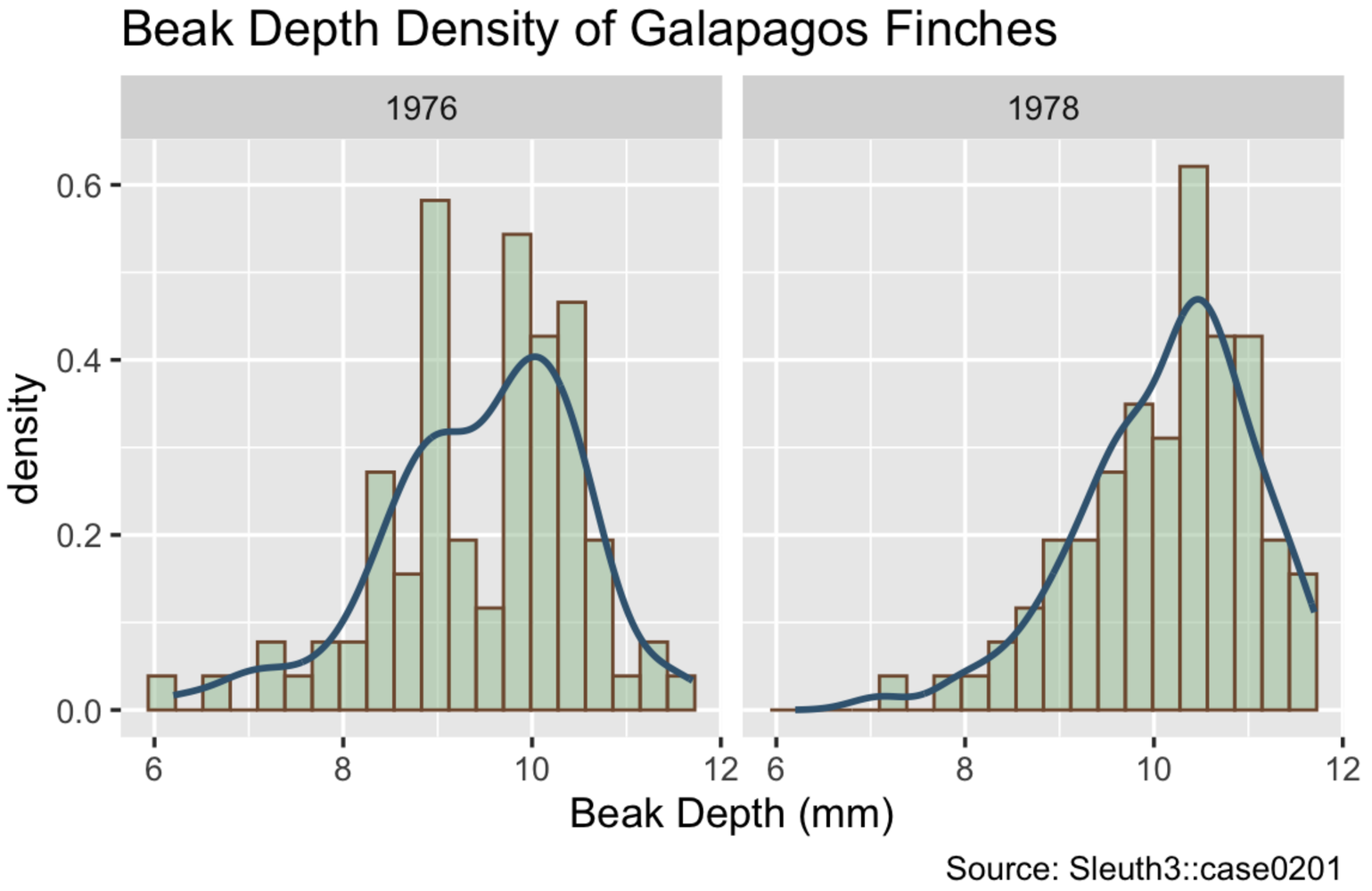
Complete the template below to build a graph.

```
ggplot (data = <DATA>) +
<GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),
 stat =<STAT>, position = <POSITION> ) +
<COORDINATE_FUNCTION>+
<FACET_FUNCTION>+
<SCALE_FUNCTION>+
 labs() +
<THEME_FUNCTION>
```

required

Not required, sensible defaults supplied

Example


Beak Depth Density of Galapagos Finches
1976
1978
density
0.6
0.4
0.2
0.0
6
8
10
12
Beak Depth (mm)
Source: Sleuth3::case0201