

Grammar of Graphics

Data layers, two mappings

`slides/03_datalayer2.pdf`

Some (fake) data

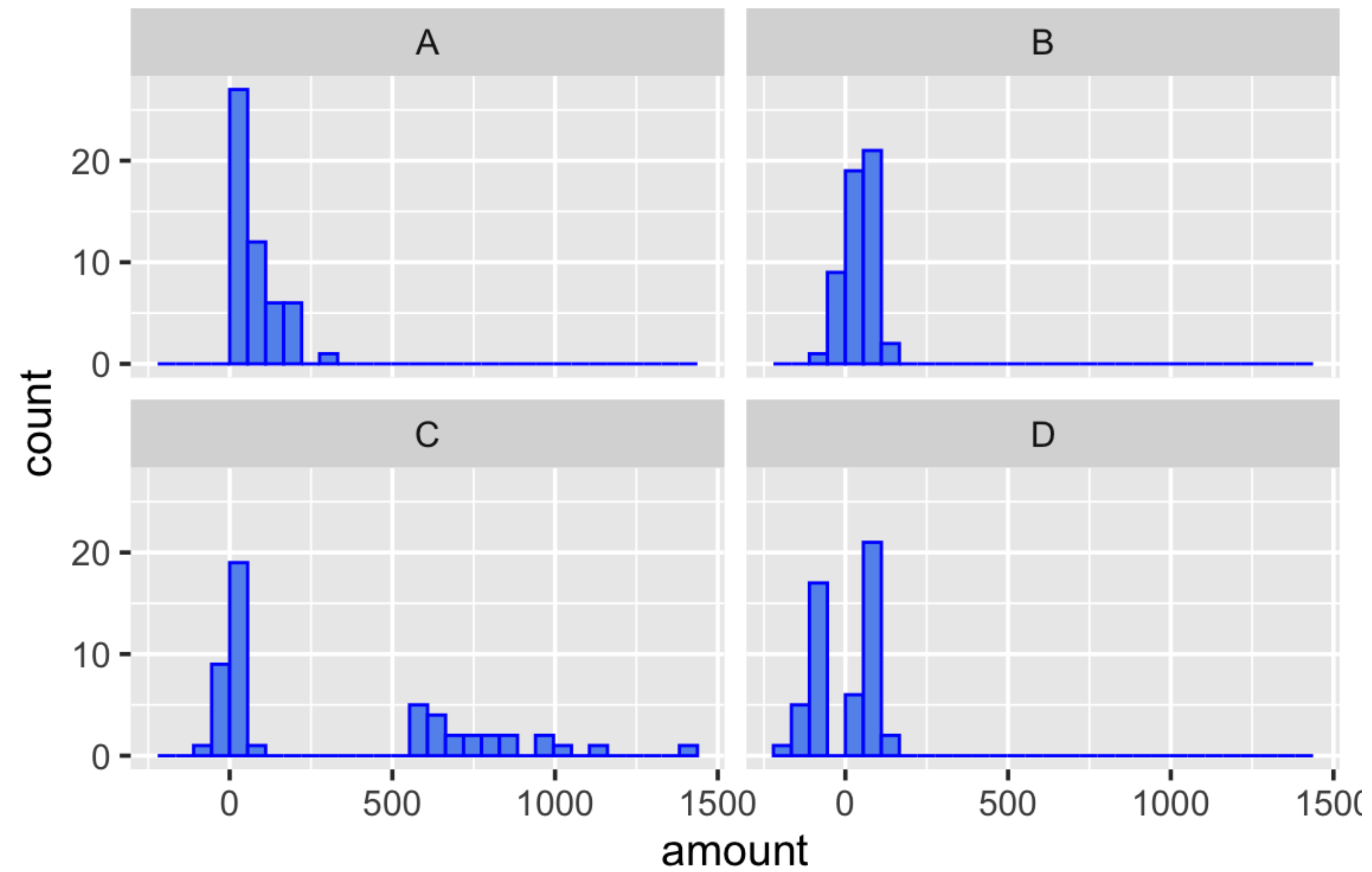
```
dim(savings)
```

```
[1] 52  4
```

```
head(savings, 10)
```

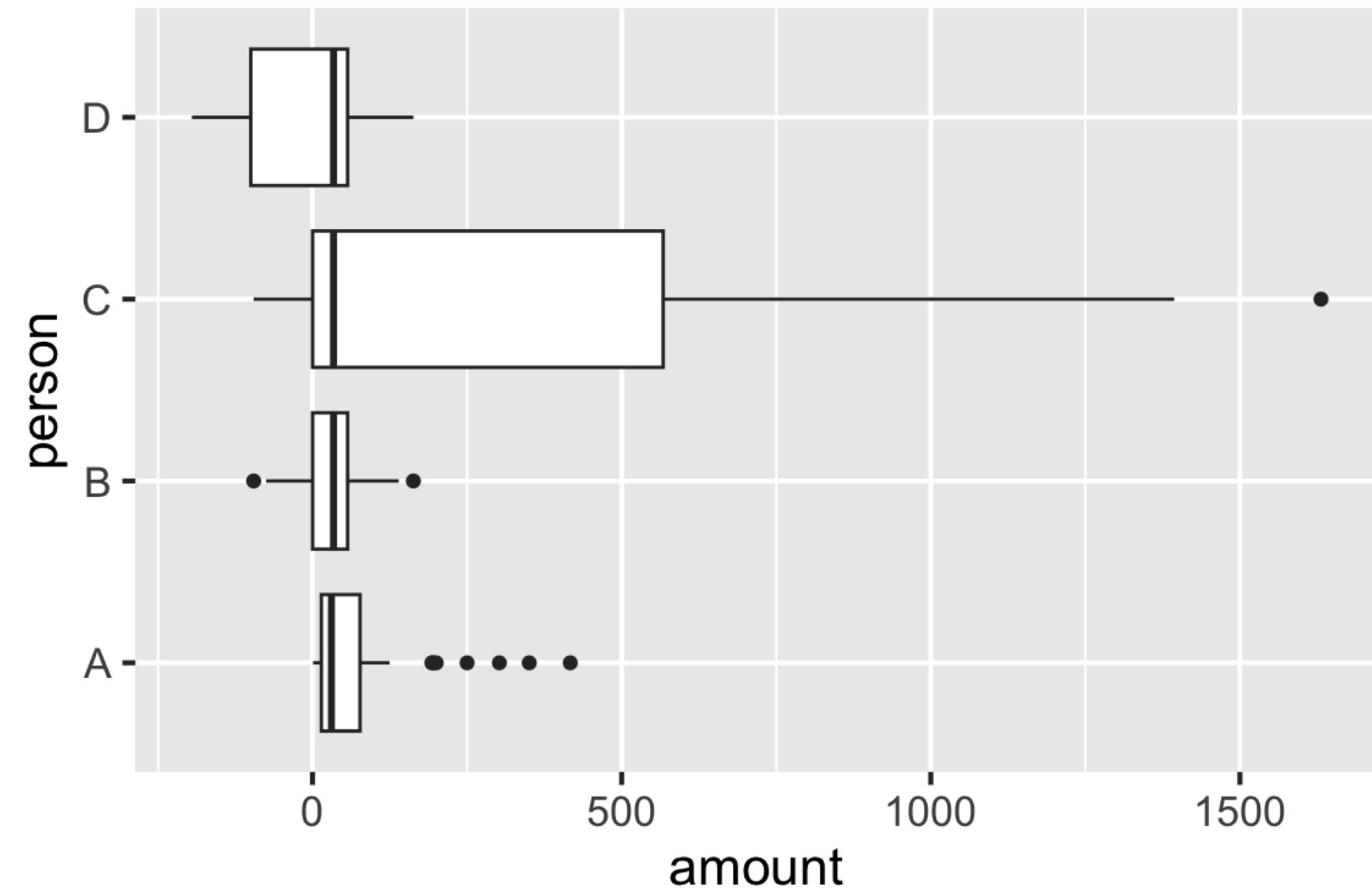
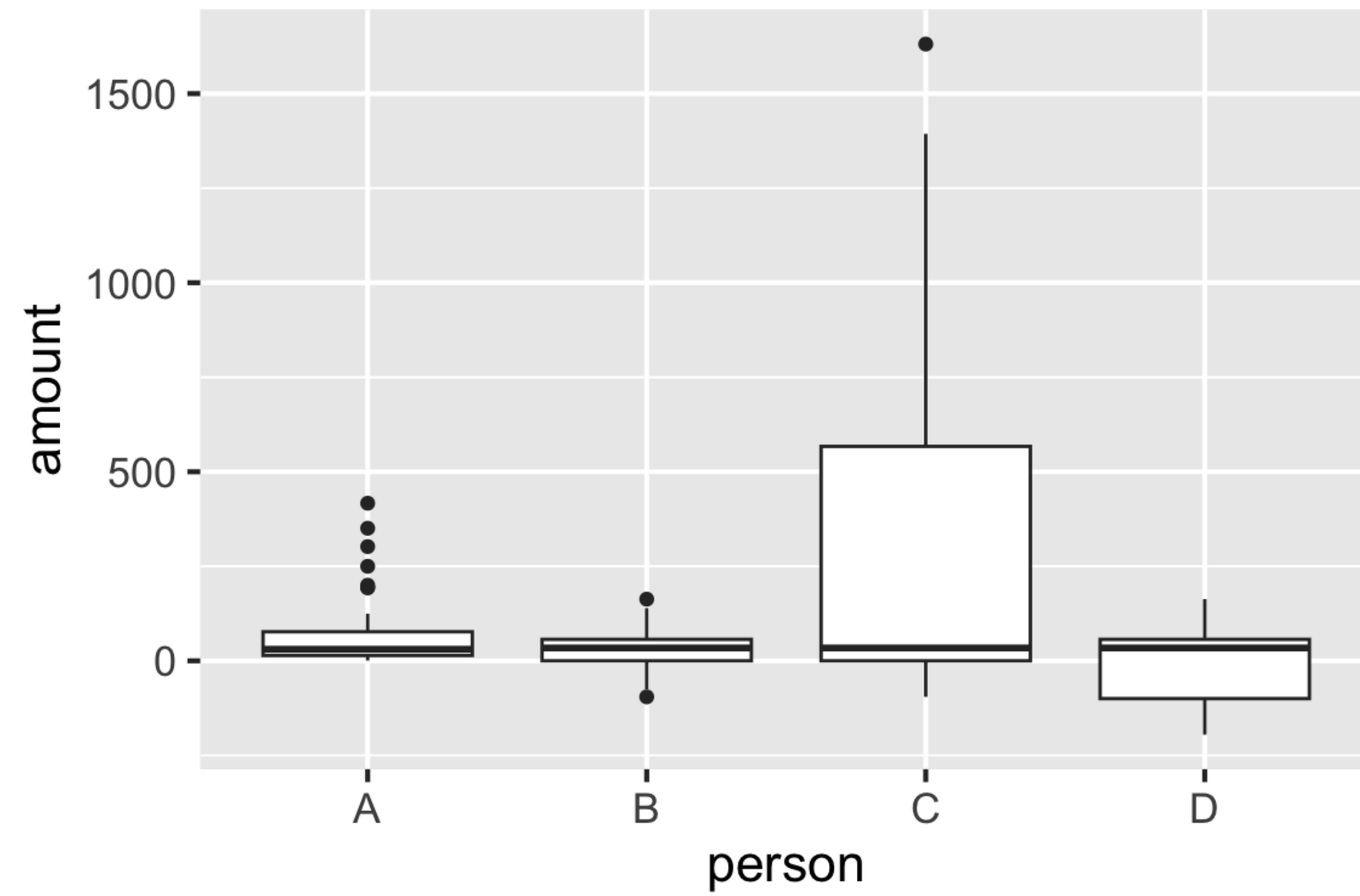
	A	B	C	D
1	0.05357402	-107.140634	-107.140634	-207.1406
2	1.09880924	-38.648645	-38.648645	-138.6486
3	1.40315630	-36.121526	-36.121526	-136.1215
4	2.73317117	-33.121797	-33.121797	-133.1218
5	3.88651907	-29.289744	-29.289744	-129.2897
6	4.53094445	-18.339285	-18.339285	-118.3393
7	5.33833355	-6.190043	-6.190043	-106.1900
8	5.71858728	-5.780576	-5.780576	-105.7806
9	5.87052914	-4.040111	-4.040111	-104.0401
10	6.67419013	-1.031474	-1.031474	-101.0315

Histograms

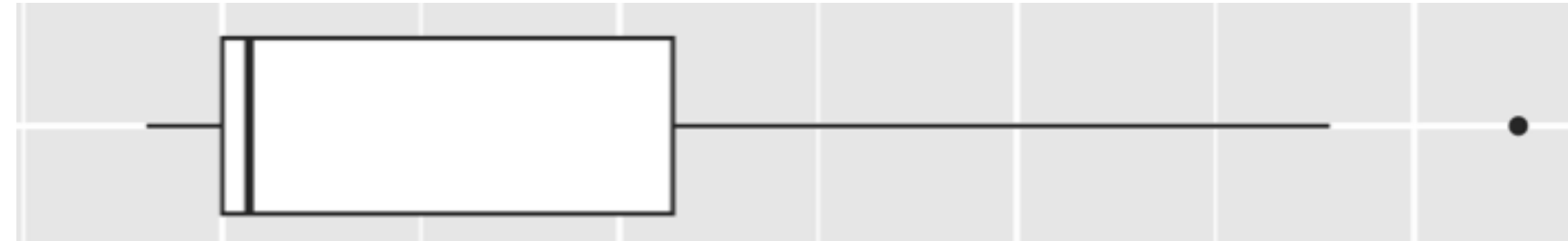


hard to compare

Boxplots



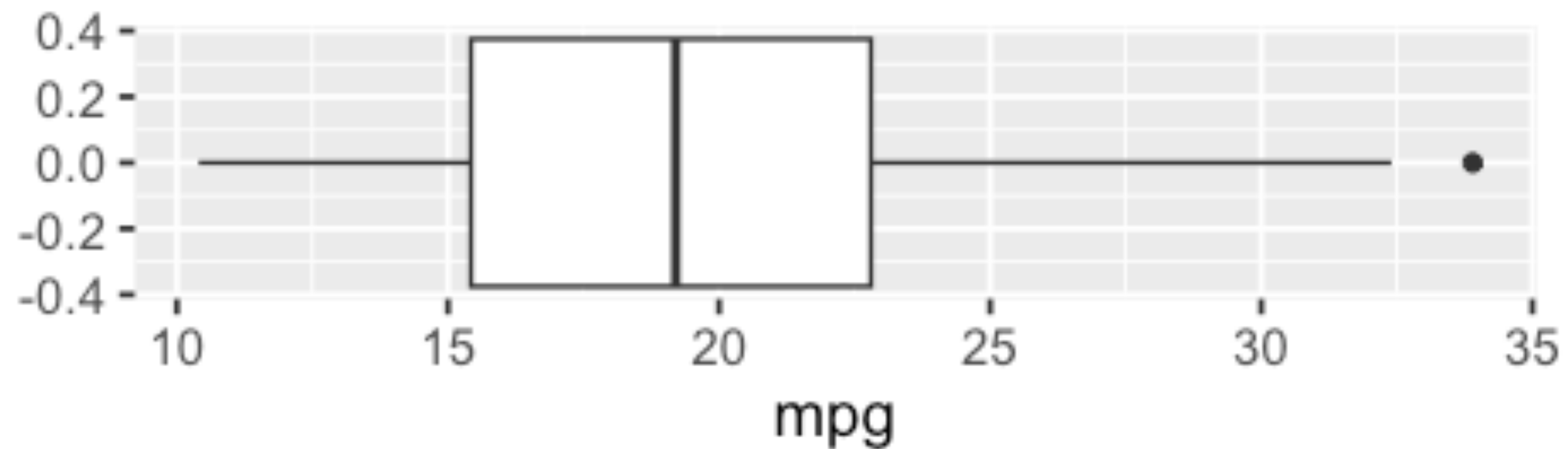
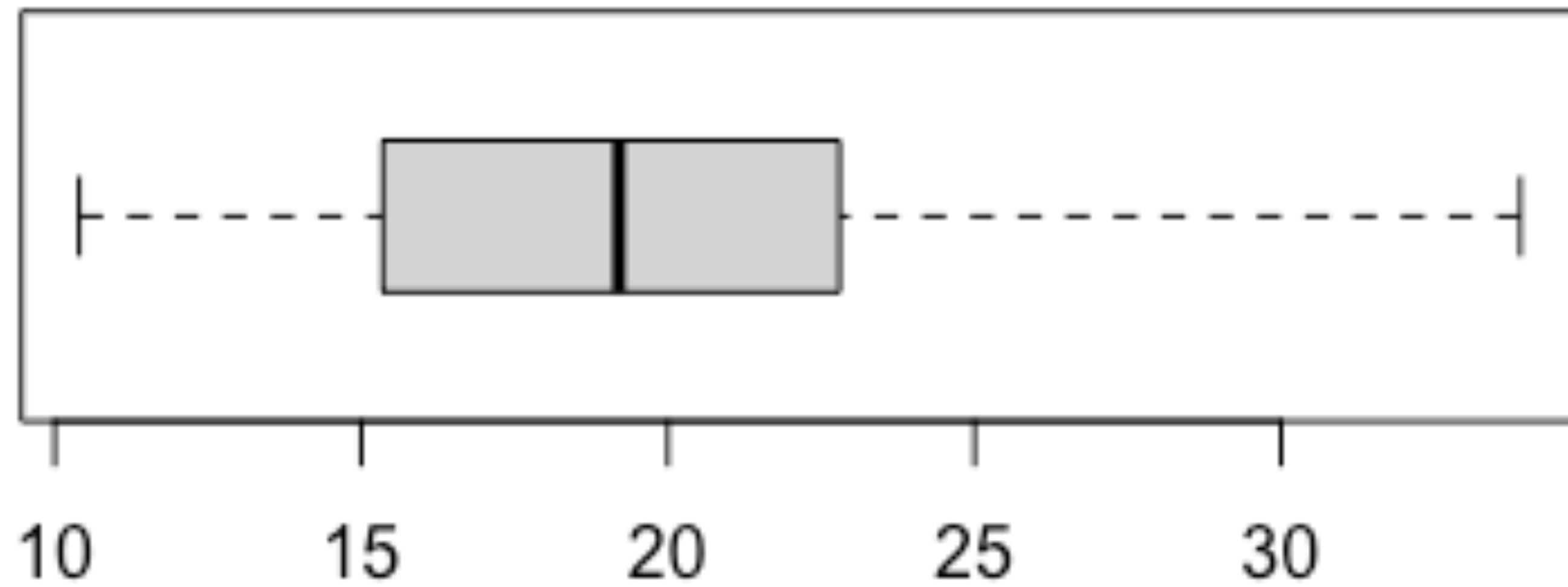
Boxplot refresher



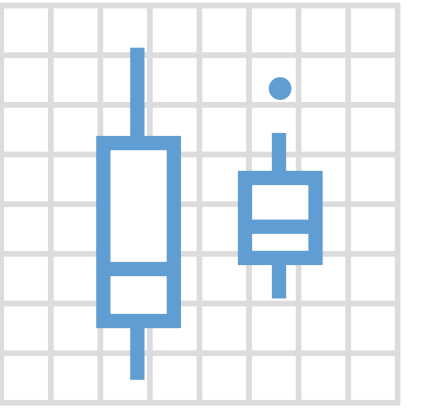
- The box represents approximately 50% of the data values
- Tukey/base R method: lower/upper hinges are medians of lower/upper halves of data
- Outliers are more than 1.5 x hinge spread above the upper hinge or 1.5 x hinge spread below the lower hinge
- ggplot2 method: use Q1, Q3, and IQR instead of lower hinge, upper hinge and hinge spread

base R vs. ggplot2

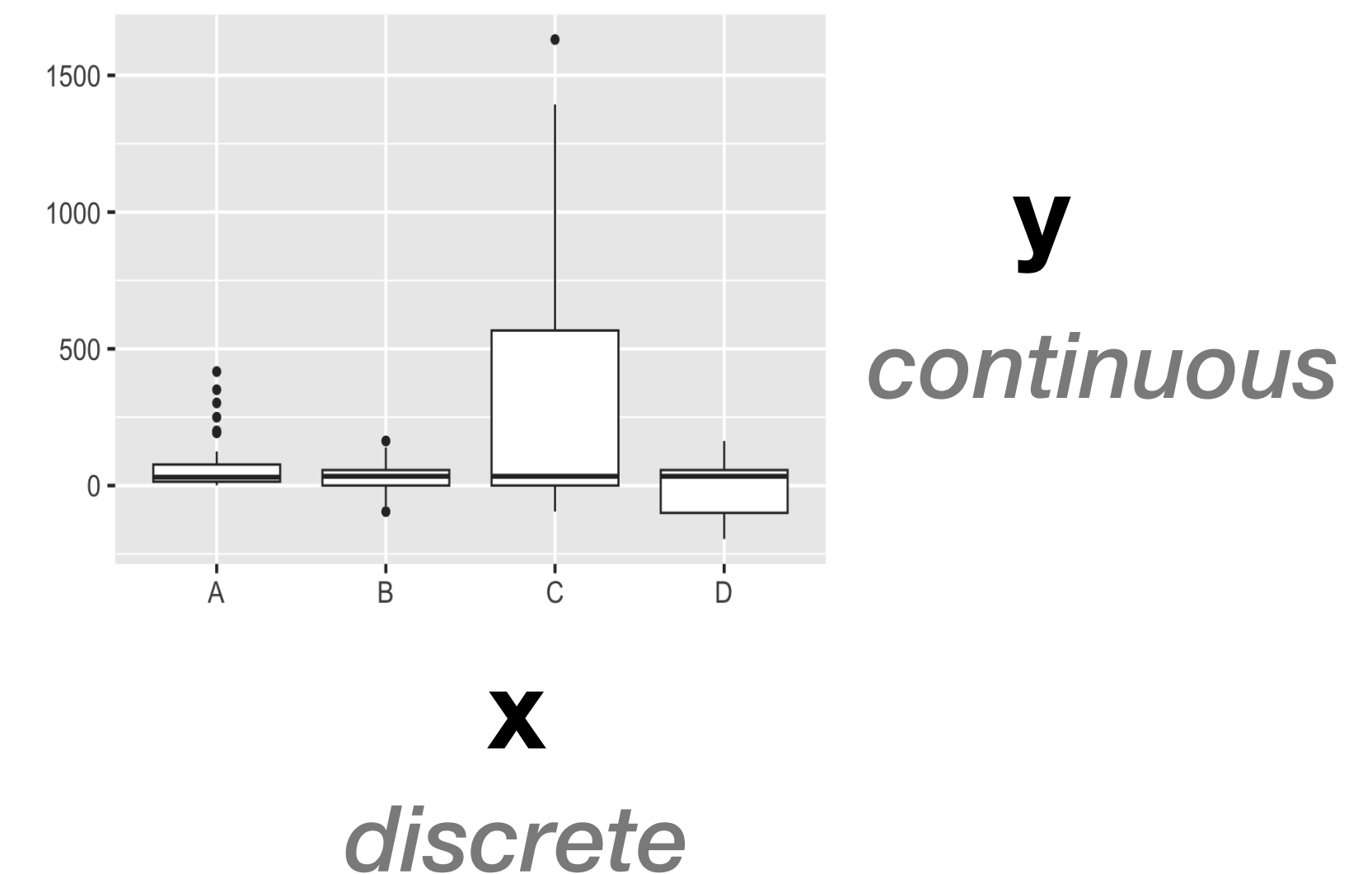
mtcars\$mpg



geom_boxplot()

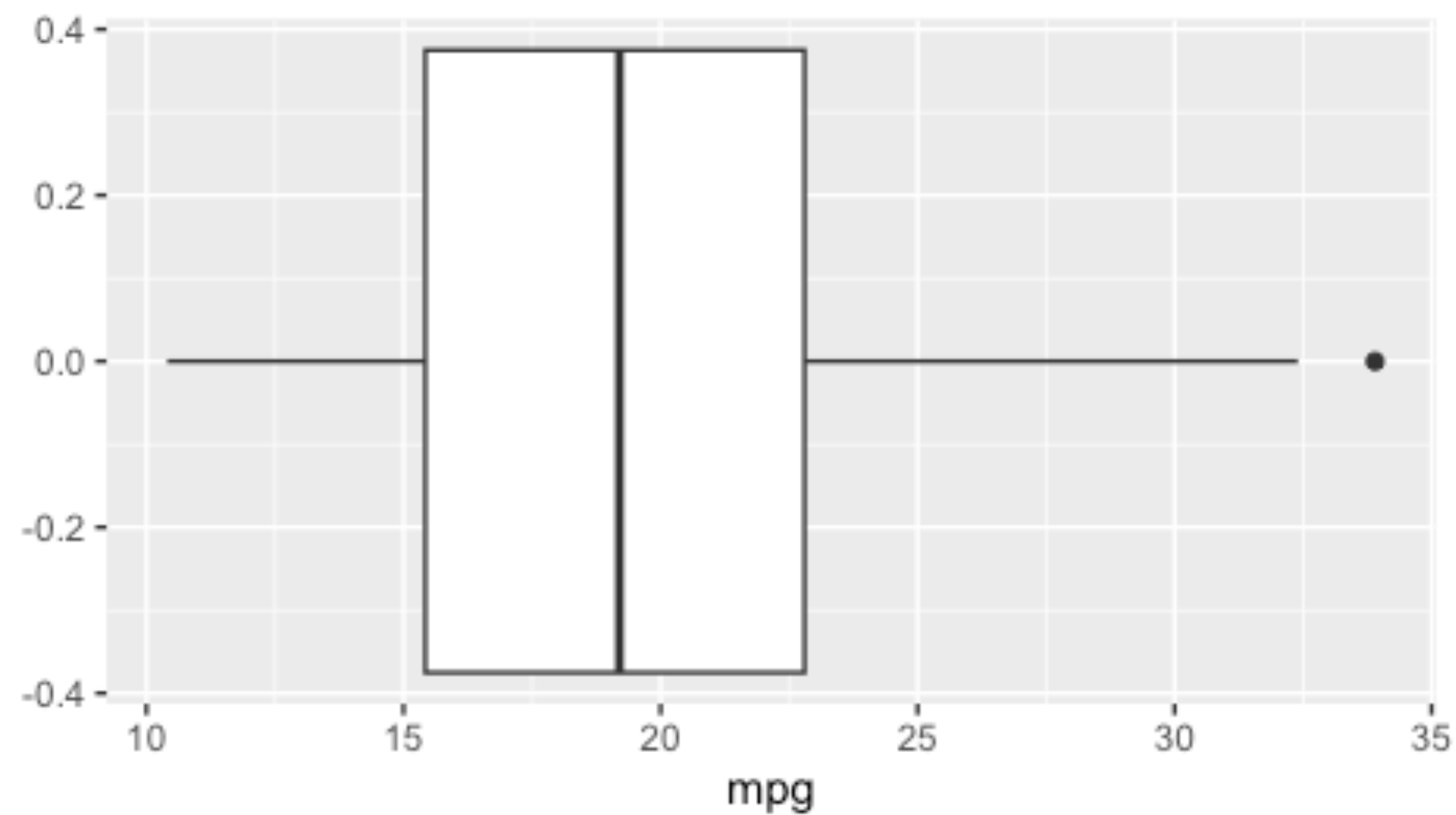


- Shows the distribution of a continuous variable
- Requires a continuous **x** (horizontal boxplots) *or y* (vertical boxplots)
- Optional discrete variable for groups (highly recommended)
- If the discrete variable is an integer, it must be converted to factor or character

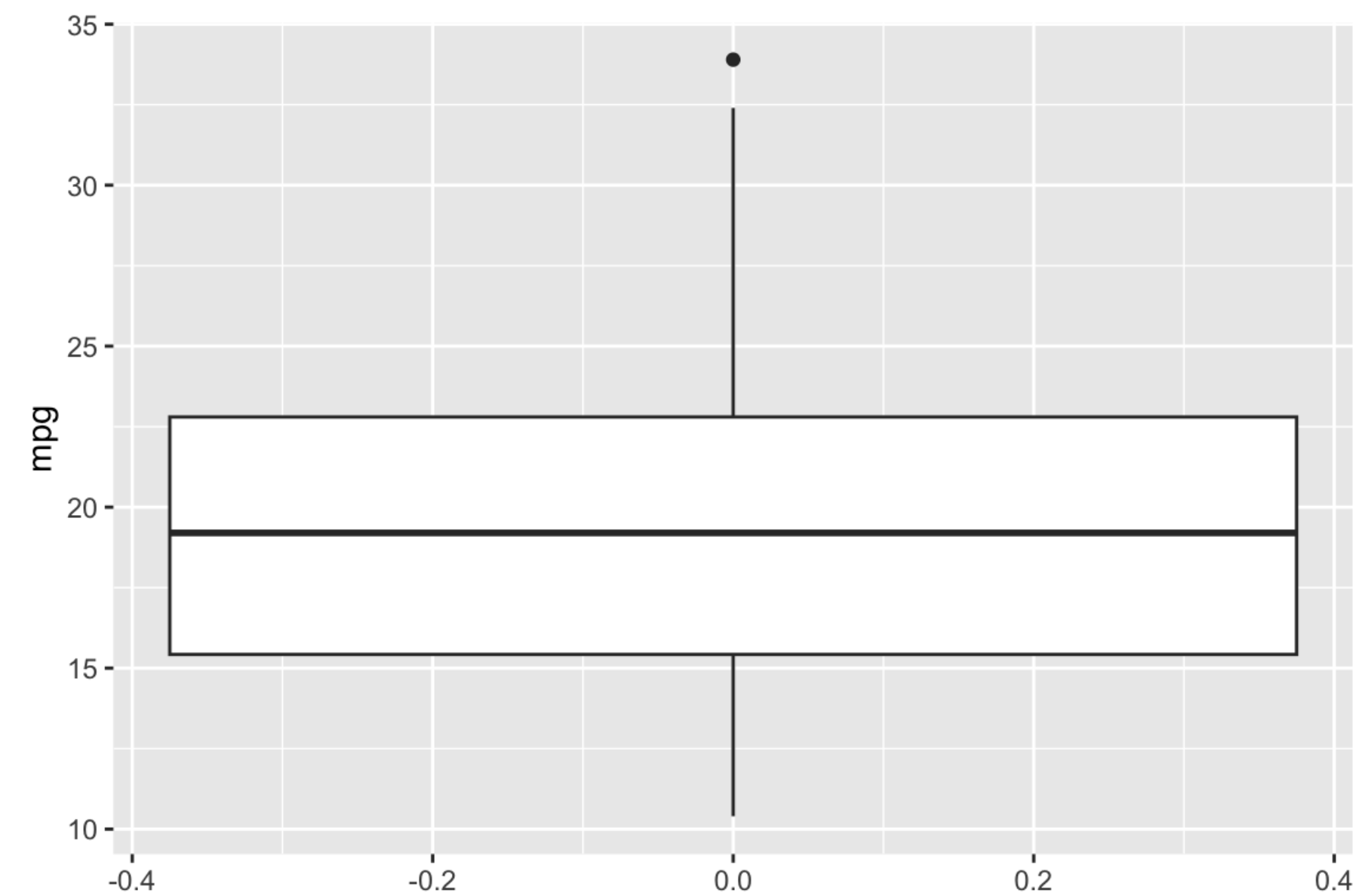


Single boxplots

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



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



The data

```
head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Look at the data

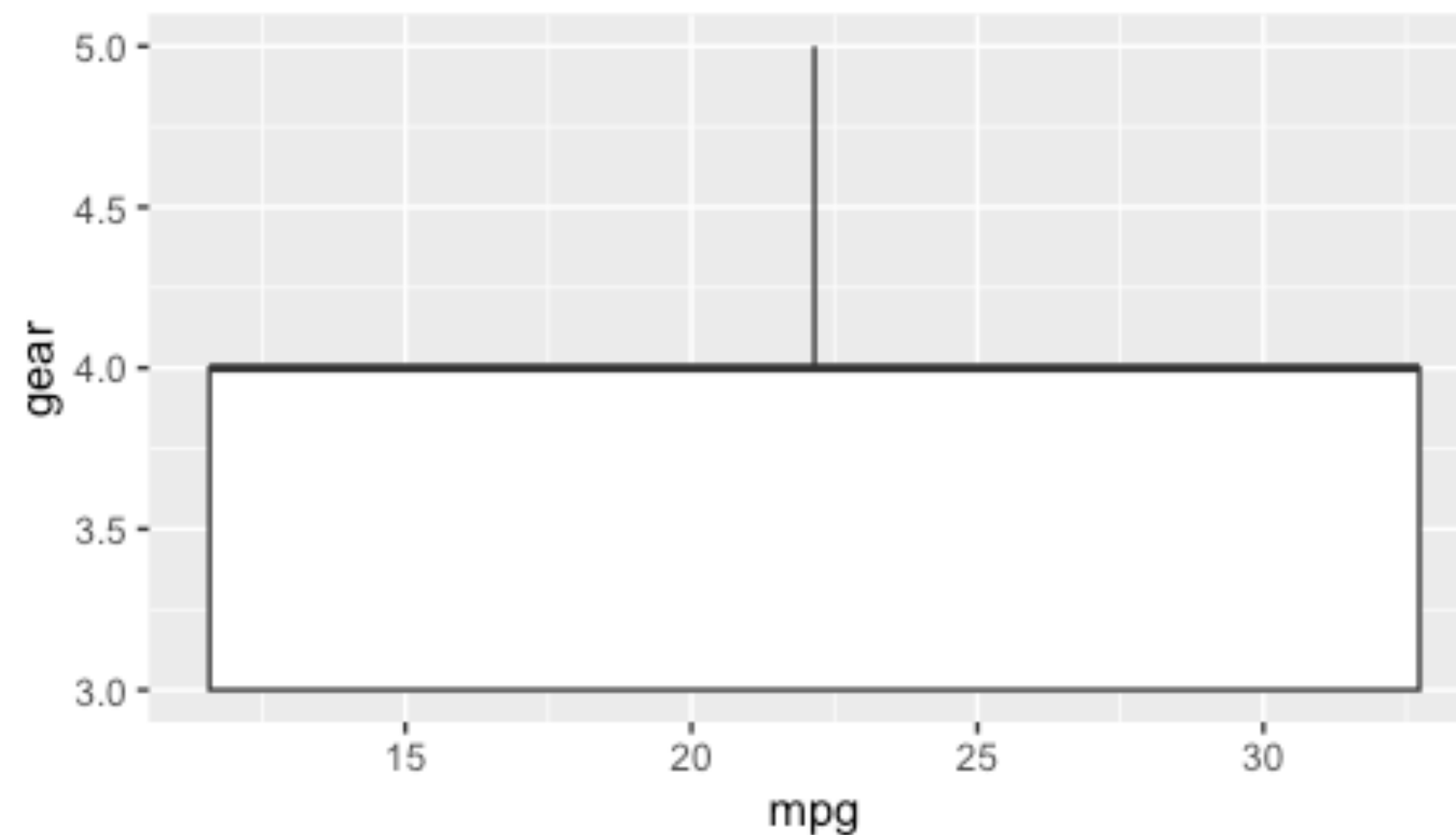
```
str(mtcars)
```

```
'data.frame':   32 obs. of  11 variables:
 $ mpg  : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl  : num   6  6  4  6  8  6  8  4  4  6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp   : num  110 110  93 110 175 105 245  62  95 123 ...
 $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt   : num   2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs   : num   0  0  1  1  0  1  0  1  1  1 ...
 $ am   : num   1  1  1  0  0  0  0  0  0  0 ...
 $ gear: num   4  4  4  3  3  3  3  4  4  4 ...
 $ carb: num   4  4  1  1  2  1  4  2  2  4 ...
```

Multiple boxplots

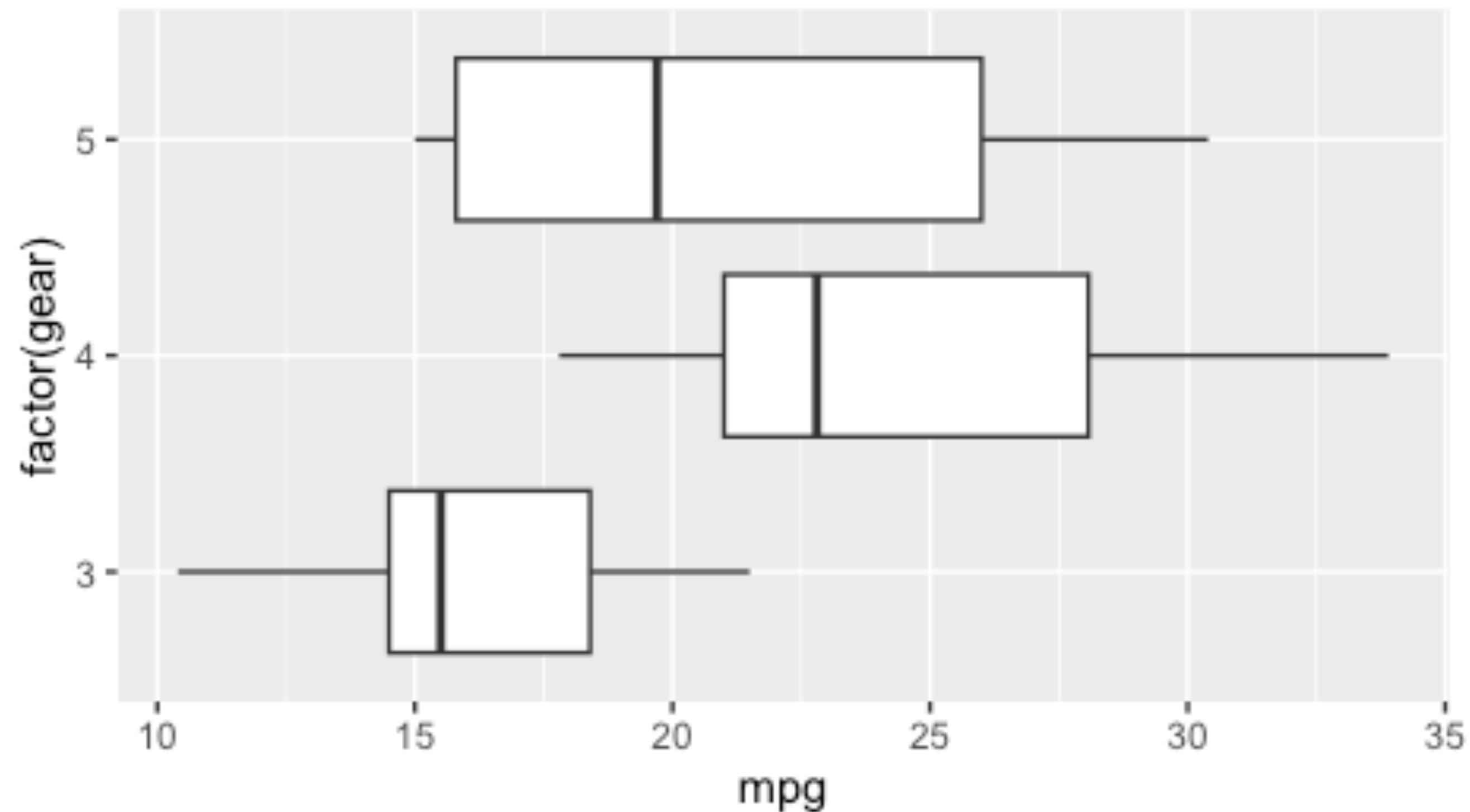
```
ggplot(mtcars, aes(x = mpg, y = gear)) +  
  geom_boxplot()
```

Warning: Continuous x aesthetic
i did you forget `aes(group = ...)`?



Multiple boxplots

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



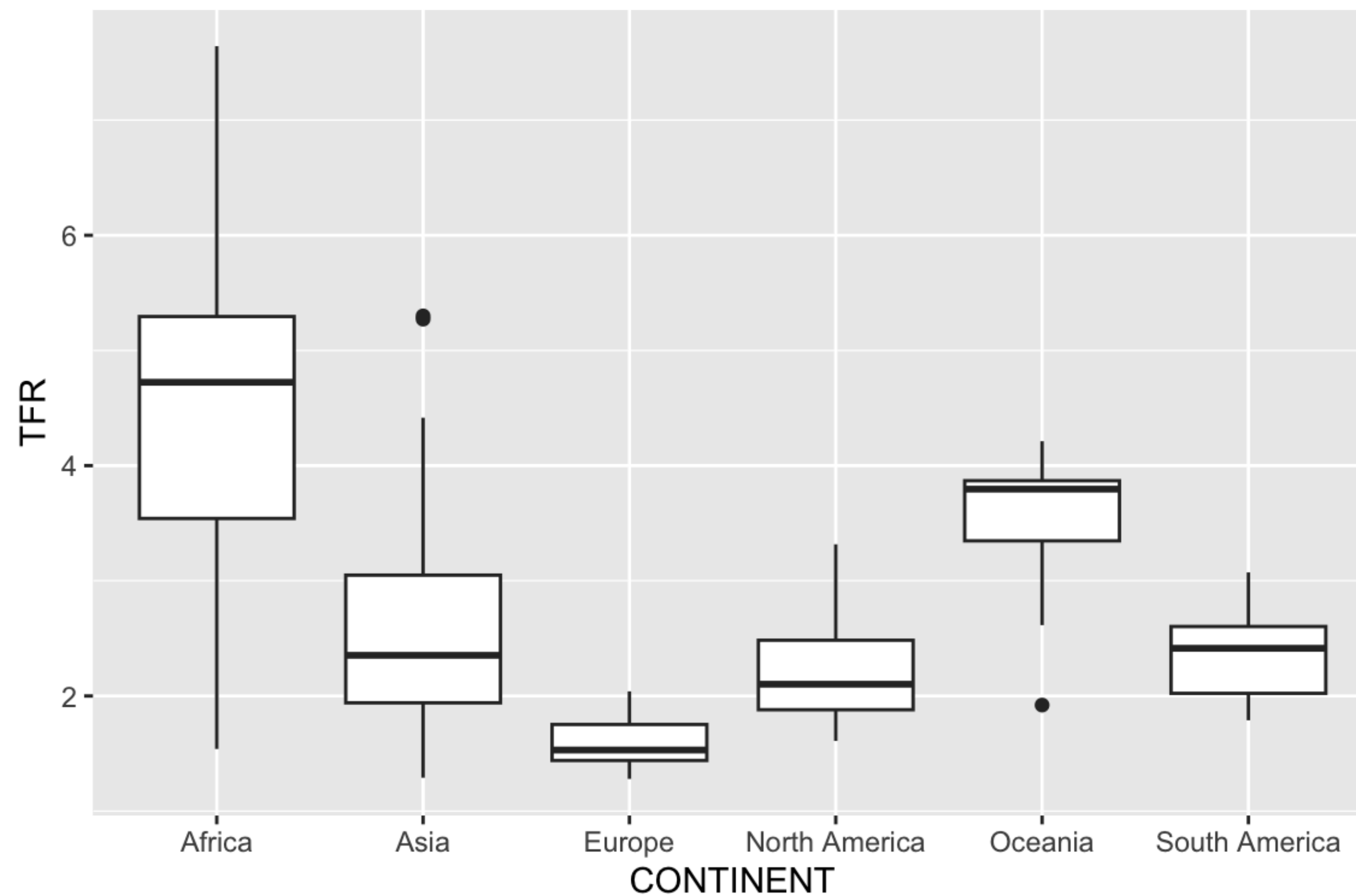
World data

```
world <- read.csv("countries2012.csv")
str(world)
```

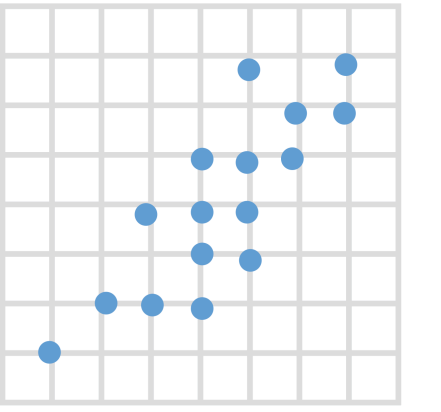
```
'data.frame':   179 obs. of  6 variables:
 $ COUNTRY   : chr  "Afghanistan" "Albania" "Algeria" "Angola" ...
 $ CONTINENT : chr  "Asia" "Europe" "Africa" "Africa" ...
 $ GDP       : num  691 4247 5584 5532 13526 ...
 $ TFR       : num  5.27 1.76 2.91 6.25 2.1 ...
 $ LIFEEXP   : num  59.7 77.4 74.3 51.5 75.6 ...
 $ CHMORT    : num  99.5 15.5 26.1 172.2 9.1 ...
```

Multiple boxplots

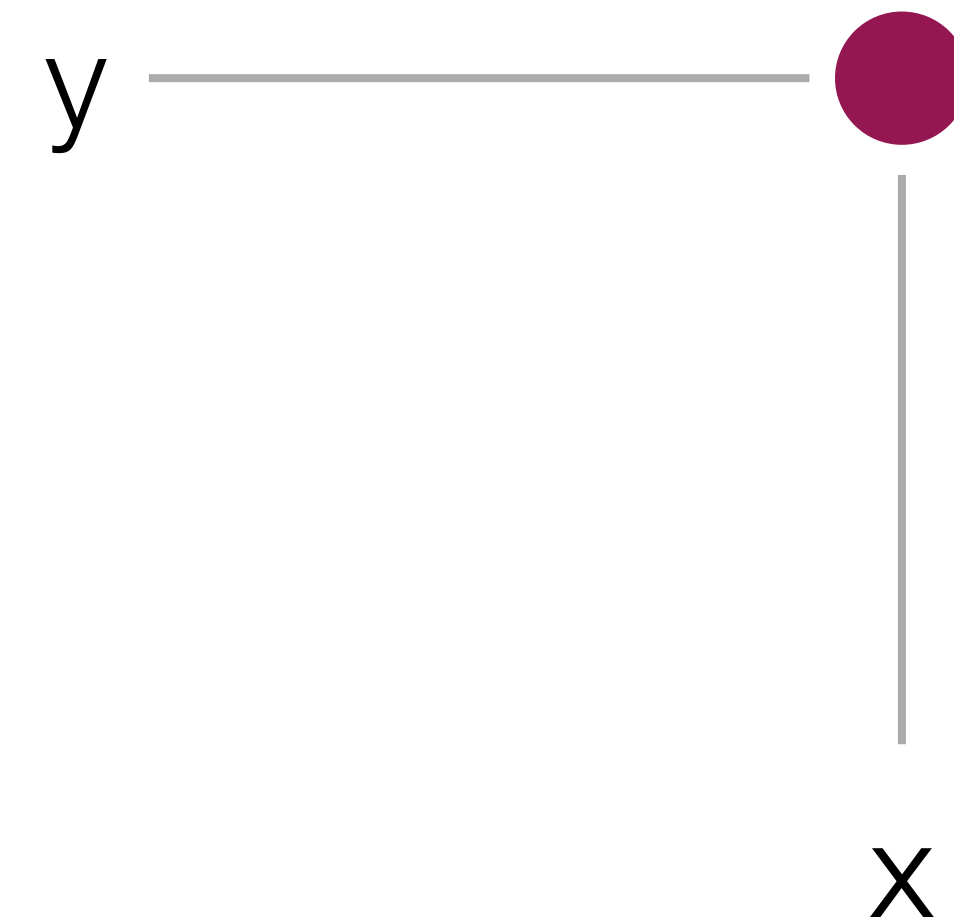
```
ggplot(world, aes(x = CONTINENT, y = TFR)) +  
  geom_boxplot()
```



geom_point()

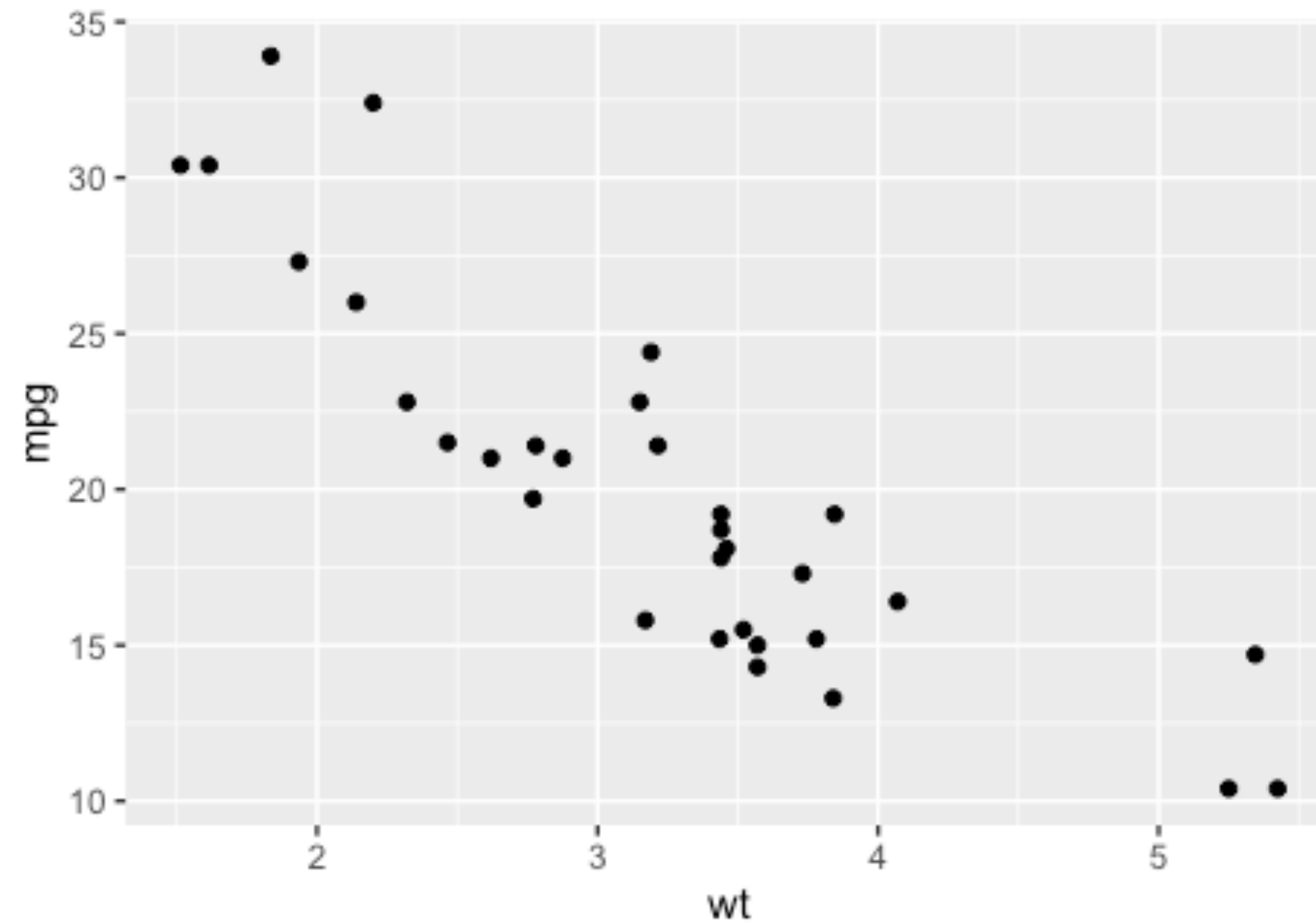


- Requires an **x** and **y**
- Both can be either continuous or discrete
- Both continuous: scatterplot
- One continuous, one discrete
Cleveland dot plot



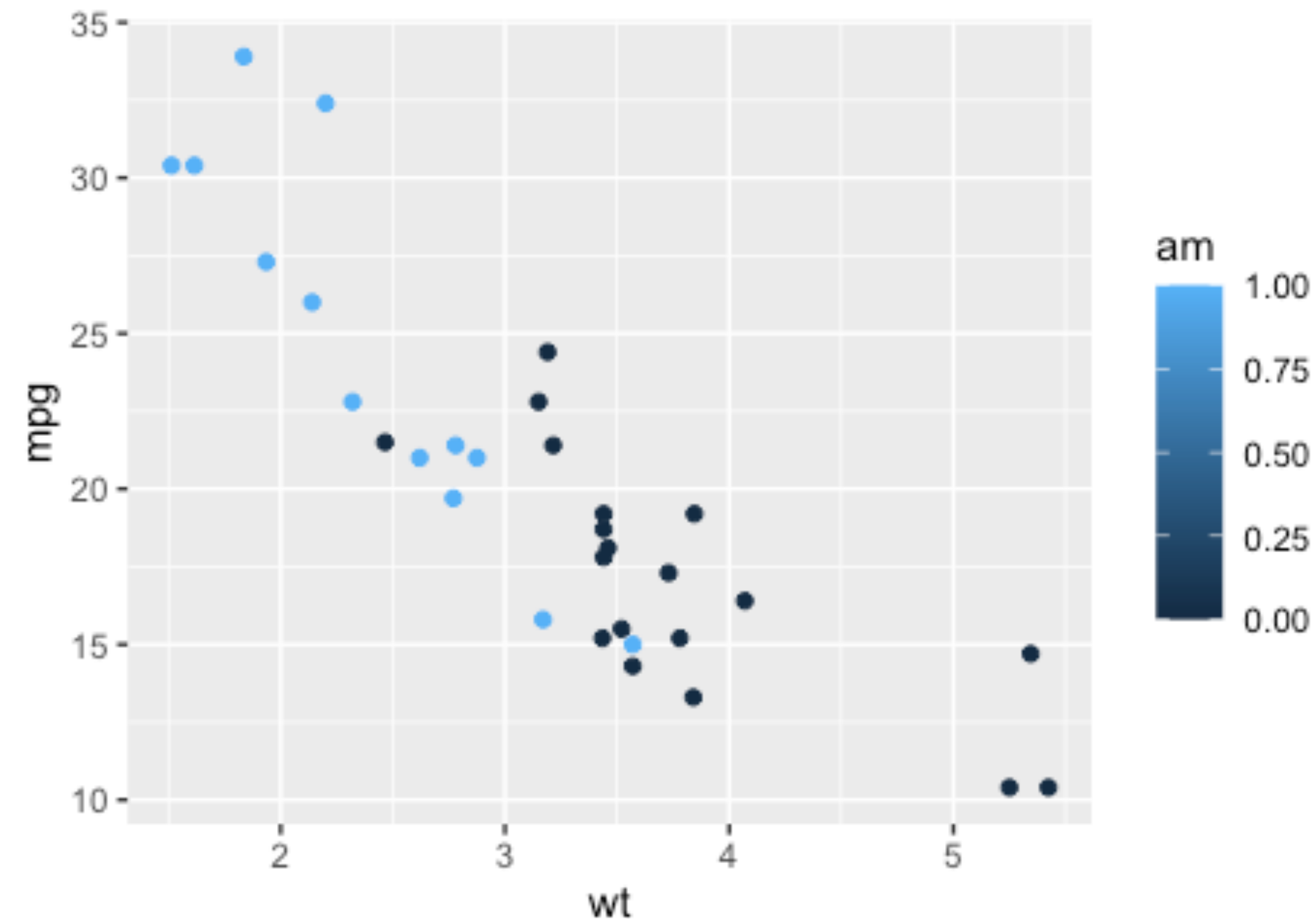
Scatterplot

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



Scatterplot with color

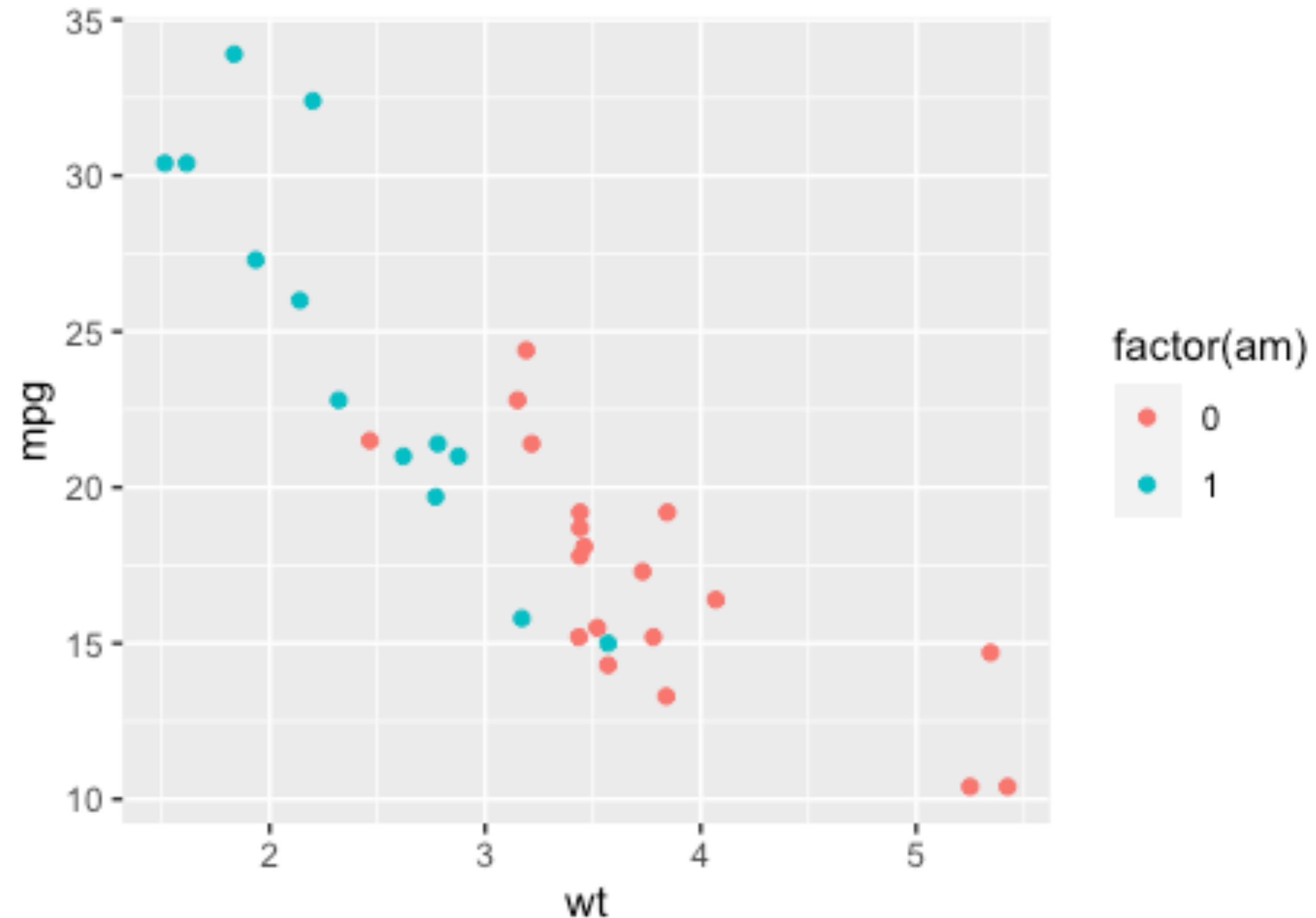
```
ggplot(mtcars, aes(x = wt, y = mpg, color = am)) +  
  geom_point()
```



What's wrong?

Scatterplot with color

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



EXERCISES

- Code: www.github.com/jtr13/csp2024
- Open `geom_boxplot.Rmd` (or `.R`)
- Run the code.
- Make changes and see what happens.
- Try the exercises.
- Repeat with `geom_point.Rmd` (or `.R`)