11. Basics of ggplot2

A **ggplot2 Primer**

The layered grammar of graphics is implemented in ggplot2, a set of tools that map data to visual display elements, and that allow users to control the fine details of the plot display.

Most important aspect: ggplot2 can be used to think about the **logical structure** of the plot.

A ggplot2 graph has 2 main components (and optional terms):
- **aesthetic mappings** (aes – connections between data and plot elems.)
- **plot geometry** (geom – specifies the type of plot)
- *facets, *coordinates, *scales, *labels, *guides, etc.

ggplot2 Grammar
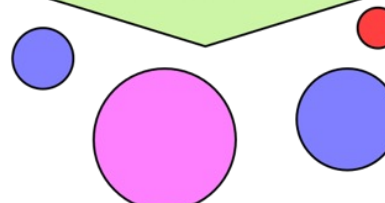
1. Tidy Data

`p ← ggplot(data = gapminder, …`

gdp	lifexp	pop	continent
340	65	31	Euro
227	51	200	Amer
909	81	80	Euro
126	40	20	Asia

2. Mapping

```
p ← ggplot(data = gapminder, mapping =
       aes(x = gdp, y = lifexp, size = pop,
           color = continent))
```
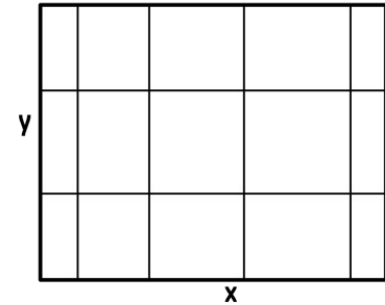
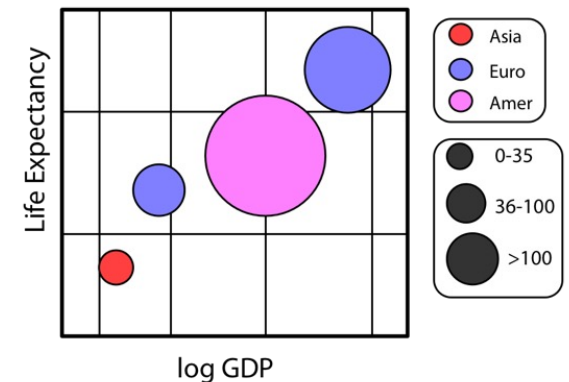## 3. Geom

`p + geom_point()`

4. Co-Ordinates & Scales

`p + coord_cartesian() + scale_x_log10()`

5. Labels & Guides

`p + labs(x = "log GDP", y = "Life Expectancy", title = "A Gapminder Plot")`

A Gapminder Plot

- Asia
- Euro
- Amer

- 0-35
- 36-100
- >100

Life Expectancy

log GDP

[Healey, K., *Data Visualization: A Practical Introduction*]

ggplot2 Grammar – Geometries

The data source and variables to be plotted are specified via ggplot().

The various geom() functions specify **how** these variables are to be visually represented:

- using points, bars, lines, shaded regions, etc.

There are currently 37+ available geometries.

Function	Adds	Options
`geom_bar()`	bar chart	color, fill, alpha
`geom_boxplot()`	boxplot	color, fill, alpha, notch, width
`geom_density()`	density plot	color, fill, alpha, linetype
`geom_histogram()`	histogram	color, fill, alpha, linetype, binwidth
`geom_hline()`	horizontal lines	color, alpha, linetype, size
`geom_line()`	jittered points	color, size, alpha, shape
`geom_jitter()`	line graph	color, alpha, linetype, size
`geom_point()`	scatterplot	color, alpha, shape, size
`geom_rug()`	rug plot	color, side
`geom_smooth()`	fitted line	method, formula, color, fill, linetype, size
`geom_text()`	text annotations	many; see the help for this function
`geom_violin()`	violin plot	color, fill, alpha, linetype
`geom_vline()`	vertical lines	color, alpha, linetype, size

Option	Specifies
color	colour of points, lines, and borders around filled regions
fill	colour of filled areas such as bars and density regions
alpha	transparency of colors, ranging from 0 (fully transparent) to 1 (opaque)
linetype	pattern for lines (1 = solid, 2 = dashed, 3 = dotted, 4 = dotdash, 5 = longdash, 6 = twodash)
size	point size and line width
shape	point shapes (same as pch, with 0 = open square, 1 = open circle, 2 = open triangle, and so on)
position	position of plotted objects such as bars and points. For bars, `dodge''` places grouped bar charts side by side, stacked" vertically stacks grouped bar charts, and `fill''` vertically stacks grouped bar charts and standardizes their heights to be equal; for points, jitter" reduces point overlap
binwidth	bin width for histograms
notch	indicates whether box plots should be notched (TRUE/FALSE)
sides	placement of rug plots on the graph (`b''` = bottom, l" = left, `t''` = top, r" = right, bl" = both bottom and left, and so on)
width	width of box plots

ggplot2 Grammar – geom()

```
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
   geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
   geom_boxplot()
```
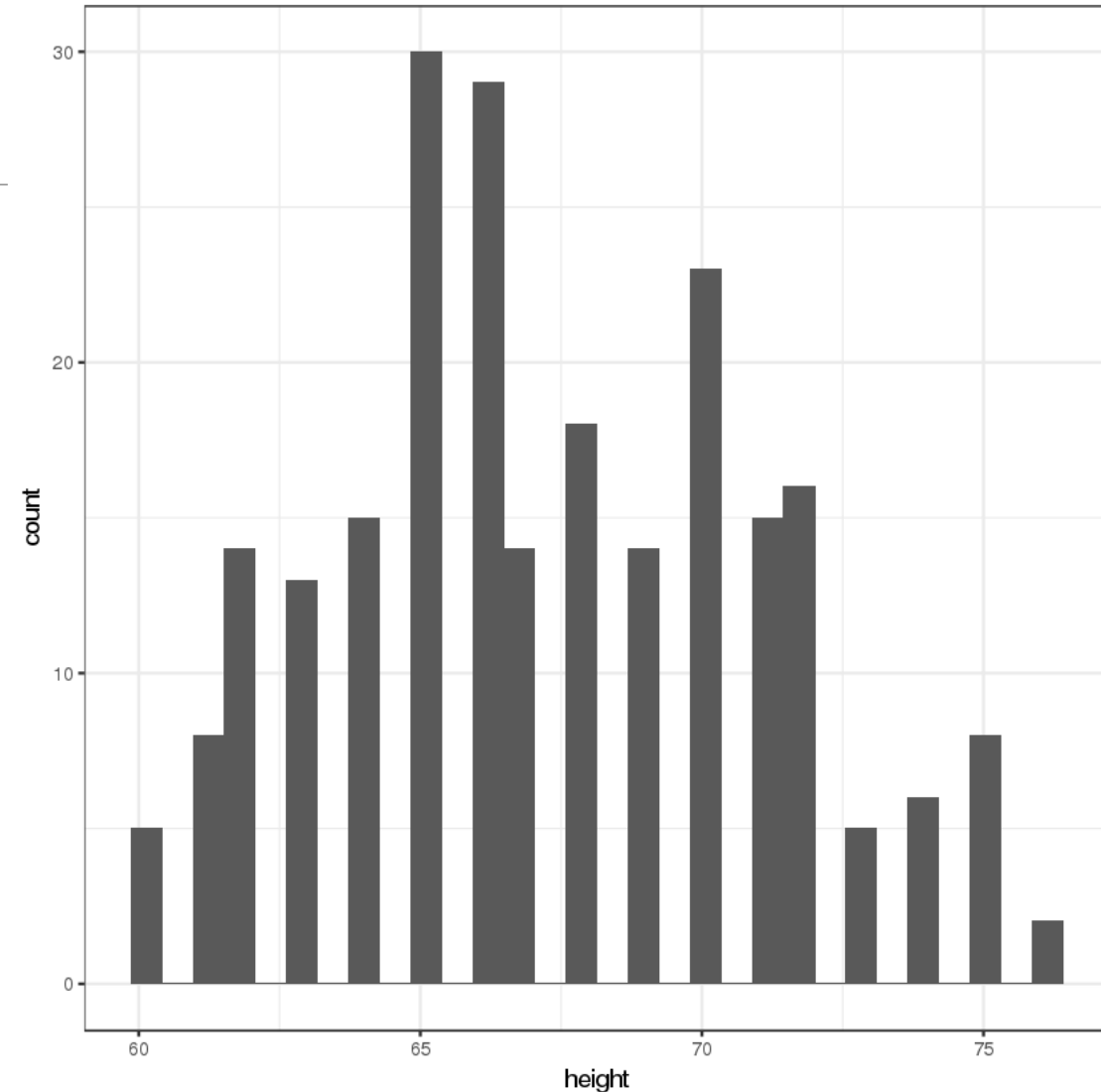
What do you expect the output to be?

ggplot2 Grammar

```
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
    geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
    geom_boxplot()
```
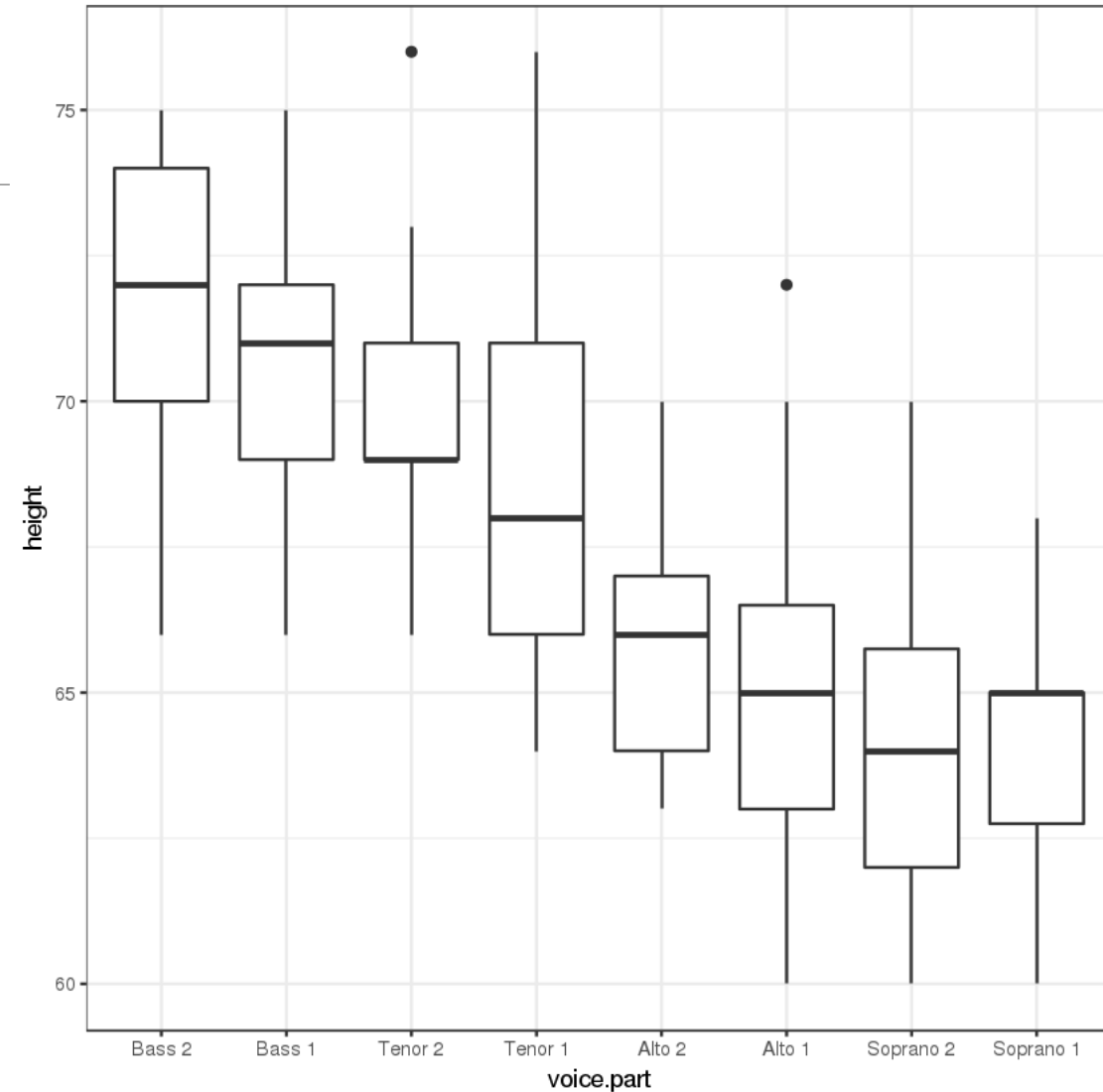
ggplot2 Grammar

```
library("ggplot2")
data(singer, package="lattice")
# Using data from the 1979 ed. of the
# New York Choral Society

# Histogram of heights
ggplot(singer, aes(x=height)) +
    geom_histogram()

# Boxplot of heights by voice part
ggplot(singer, aes(x=voice.part, y=height)) +
    geom_boxplot()
```

ggplot2 Grammar – geom()

```
library(ggplot2)
data(Salaries, package="car")
# Using data on salaries of a sample of
# US university professors (2018-2019)
# var: rank, sex, yrs.since.phd, yrs.service, salary

ggplot(Salaries, aes(x=rank, y=salary)) +
   geom_boxplot(fill="cornflowerblue",color="black", notch=TRUE) +
   geom_point(position="jitter", color="blue", alpha=.5) +
   geom_rug(side="l", color="black")
```

What do you expect the output to be?

ggplot2 Grammar


```
library(ggplot2)
data(Salaries, package="car")
# Using data on salaries of a sample of
# US university professors (2018-2019)
# var: rank, sex, yrs.since.phd, yrs.service, salary

ggplot(Salaries, aes(x=rank, y=salary)) +
    geom_boxplot(fill="cornflowerblue",color="black", notch=TRUE) +
    geom_point(position="jitter", color="blue", alpha=.5) +
    geom_rug(side="l", color="black")
```
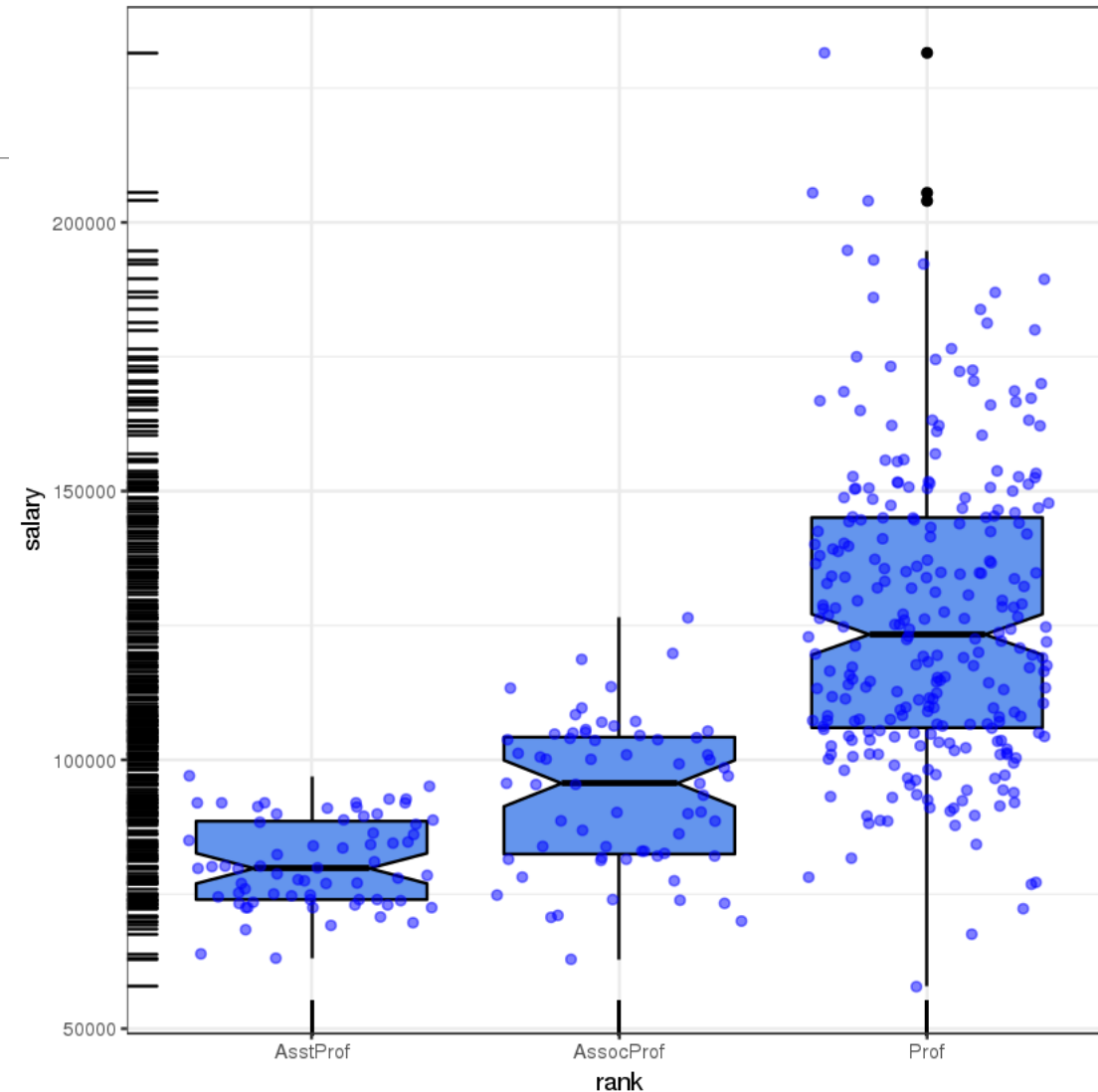
ggplot2 Grammar – Aesthetics

Aesthetics refer to the displayed attributes of the data.

They map the data to an attribute (such as the size or shape of a marker) and generate an appropriate legend.

Aesthetics are specified with the aes() function, either within the data call or within a geom() call. If they're specified within ggplot() then they apply to all specified geometries.

ggplot2 Grammar – Aesthetics

The aesthetics available to geom_point() (scatterplot), as an example, are:

- x , y , alpha , colour , fill , shape, size

Important difference between specifying characteristics (like colour, shape) inside and outside the aes() call

- **inside:** assigned colour or shape automatically based on the data
- **outside**: not mapped to data

ggplot2 Grammar

```
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
    geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
    geom_point(colour = "red")
```
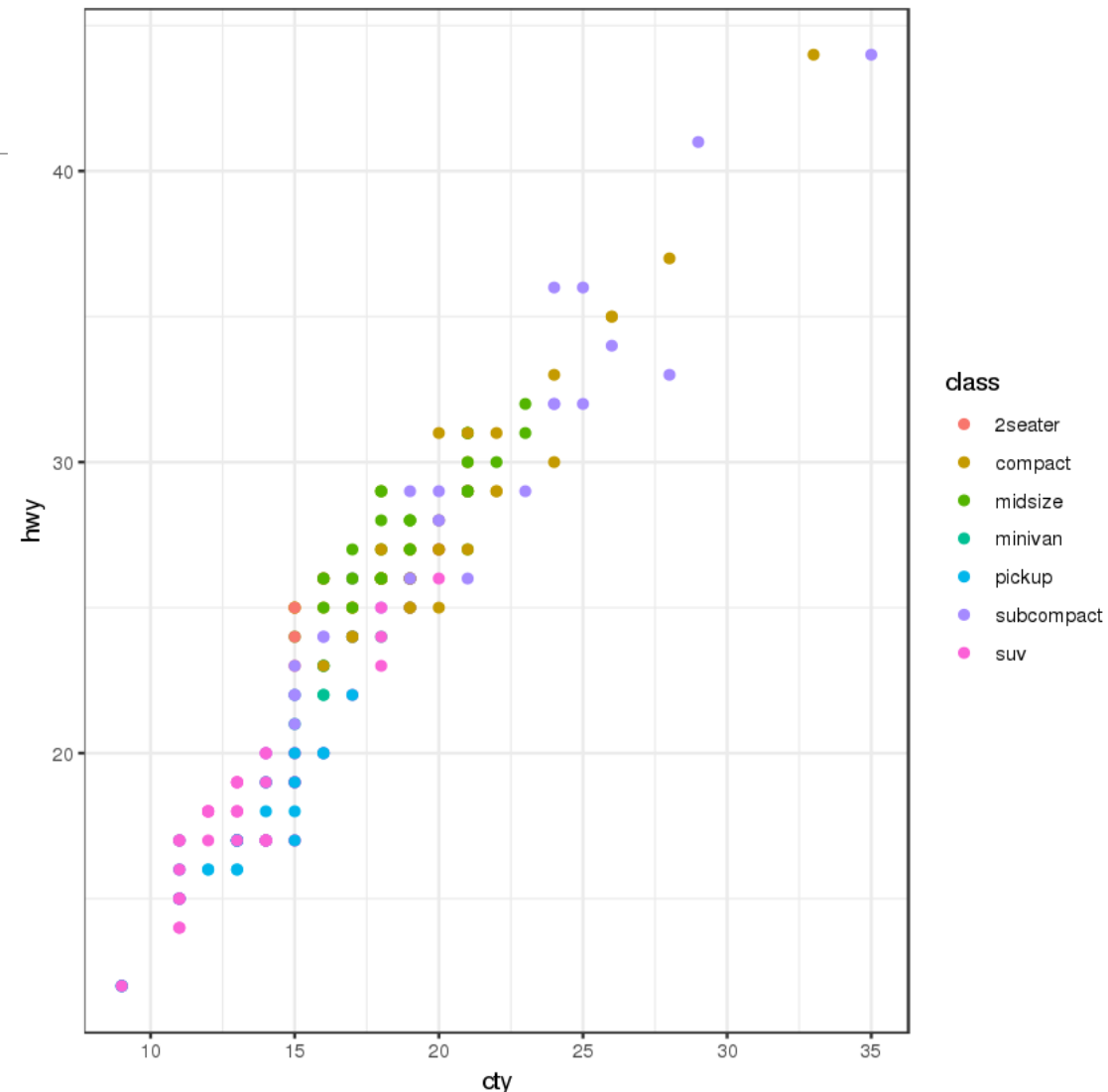
What do you expect the output to be?

ggplot2 Grammar

```
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
  geom_point(colour = "red")
```
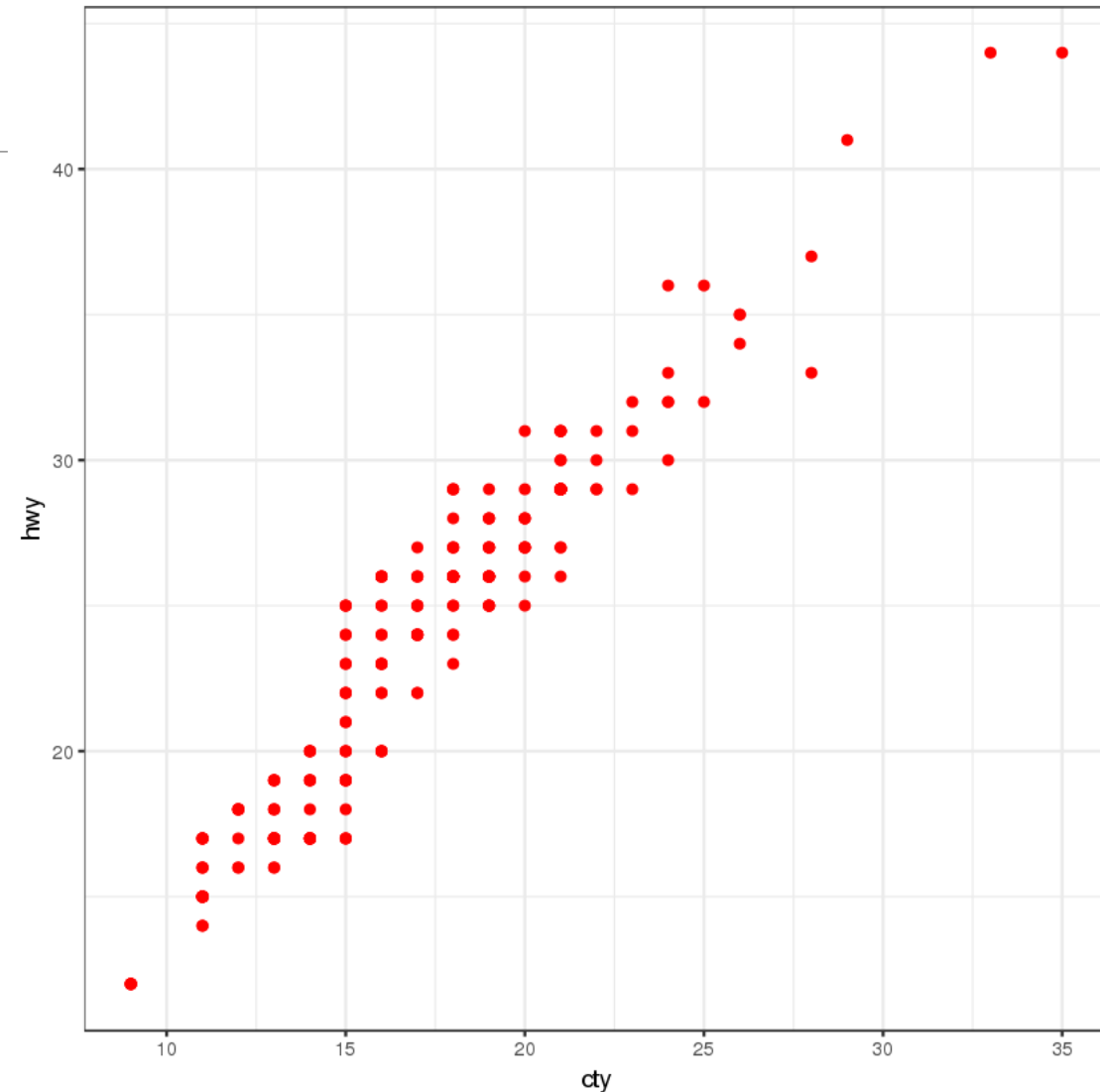
ggplot2 Grammar

```
library(ggplot2)
# Using the mpg dataset

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
    geom_point(aes(colour = class))

# specifying characteristics inside aes()
ggplot(mpg, aes(cty, hwy)) +
    geom_point(colour = "red")
```

ggplot2 Grammar – Facets

In ggplot2 parlance, small multiples are referred to as facets

- facet_wrap(), facet_grid()

By default, all panels (one for each factor) share the same axes (scale-wise).

Separating the graph into a sequence of smaller, side-by-side plots makes it easier to enact comparisons.

Wraps only display those small multiples for which there is data, grids display all multiples, even the empty ones.
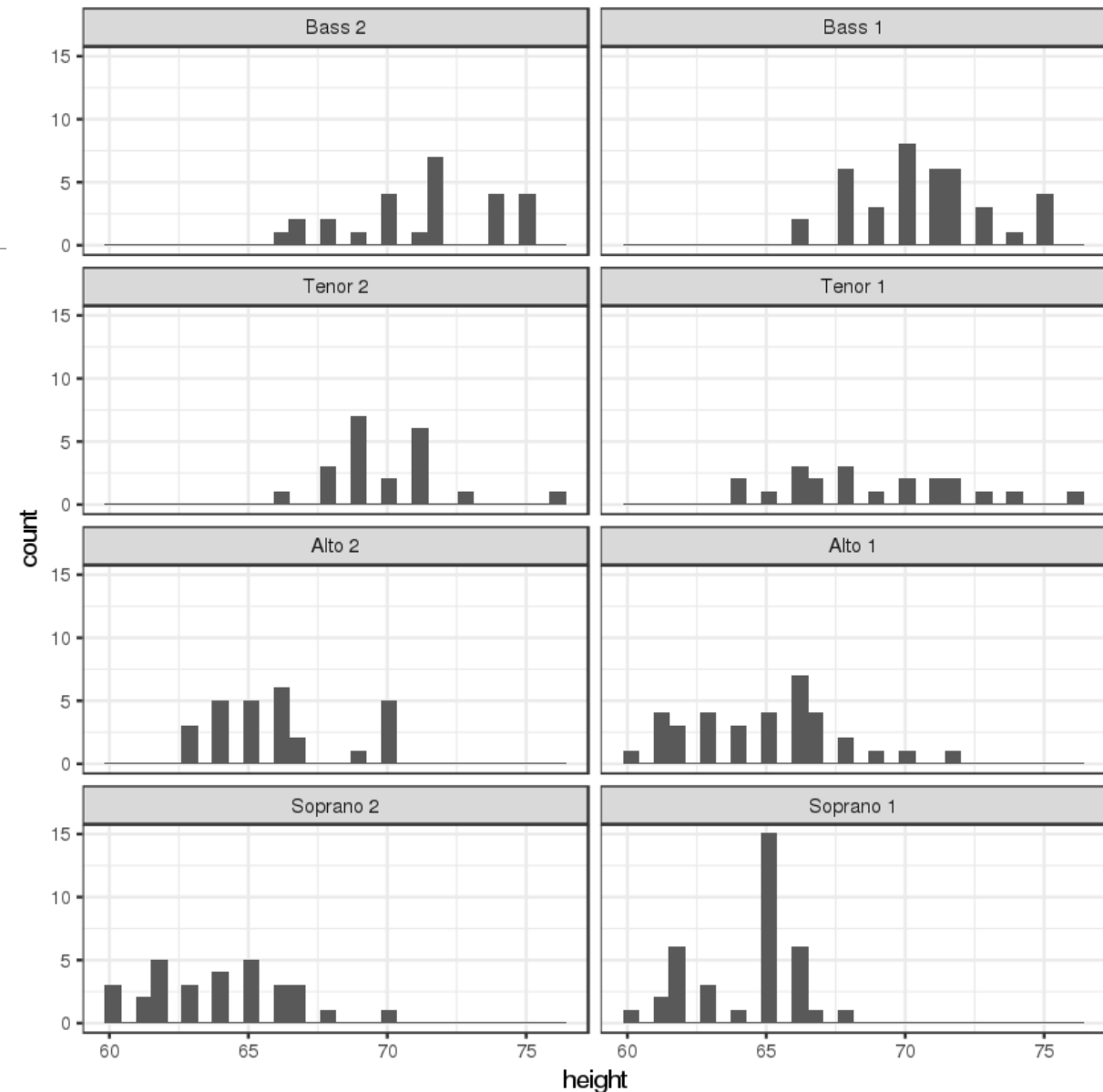
ggplot2 Grammar

```
data(singer, package="lattice")
library(ggplot2)
ggplot(data=singer, aes(x=height)) +
    geom_histogram() +
    facet_wrap(~voice.part, nrow=4)
```

What do you expect the output to be?

ggplot2 Grammar

```
data(singer, package="lattice")
library(ggplot2)
ggplot(data=singer, aes(x=height)) +
    geom_histogram() +
    facet_wrap(~voice.part, nrow=4)
```
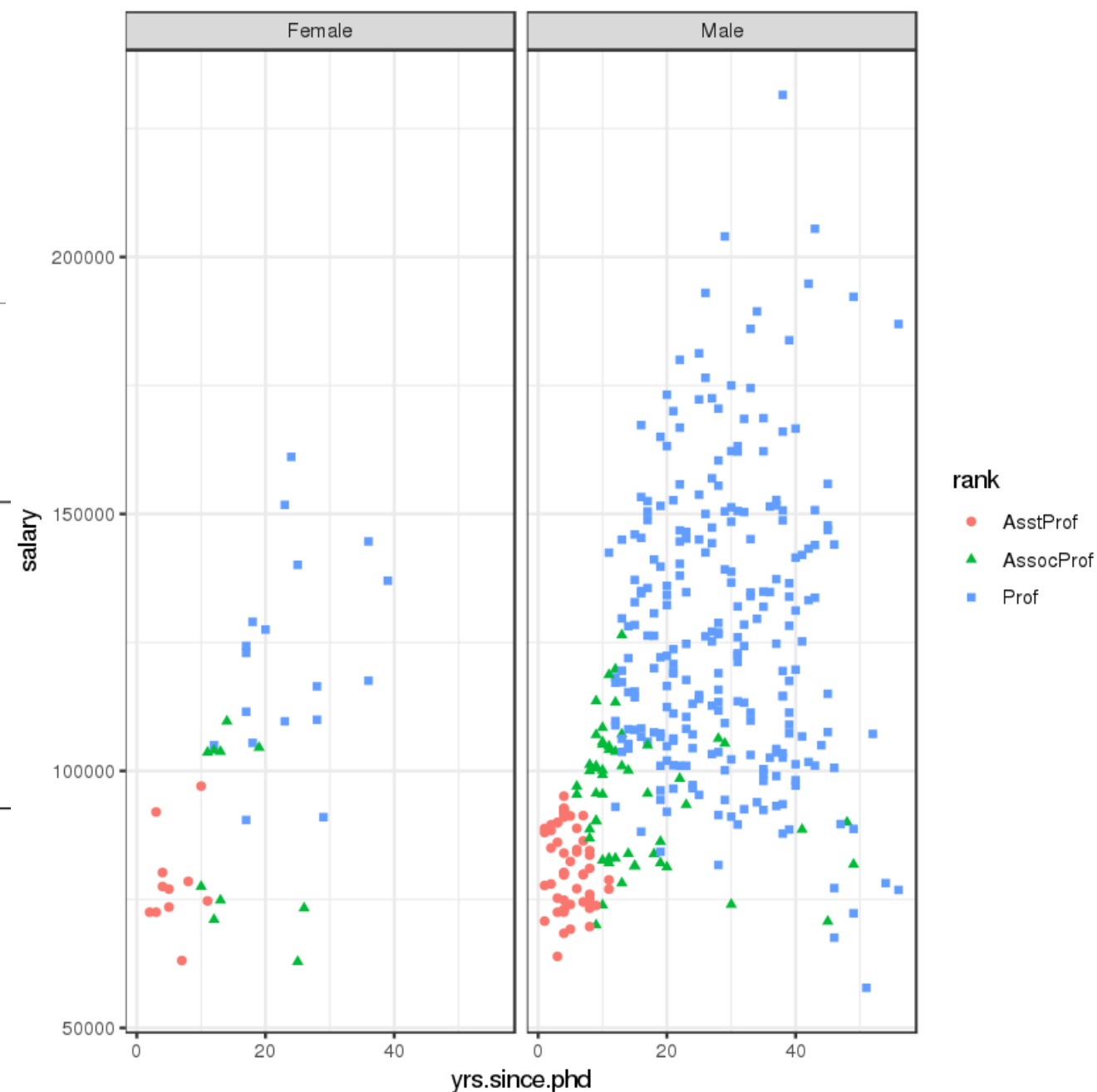
ggplot2 Grammar

```
data(Salaries, package="car")
library(ggplot2)
ggplot(Salaries, aes(x=yrs.since.phd,
 y=salary, color=rank, shape=rank)) +
   geom_point() +
   facet_grid(~sex)
```

What do you expect the output to be?

ggplot2 Grammar

```
data(Salaries, package="car")
library(ggplot2)
ggplot(Salaries, aes(x=yrs.since.phd,
 y=salary, color=rank, shape=rank)) +
   geom_point() +
   facet_grid(~sex)
```

Suggested Reading

Basics of ggplot2

Data Understanding, Data Analysis, Data Science
Data Visualization and Data Exploration

ggplot2 Visualizations in R

- Basics of ggplot2's Grammar

- ggplot2 Miscellanea

Exercises

Basics of ggplot2

Create some simple ggplot2 visualizations with data available in R. The emphasis is on becoming familiar with various geometries, their aesthetics, and the use of facets. You may use the examples found in this module as the basis of your work.

The available datasets are obtained by running

```
> data()
```

at the R command prompt.