
DATA VISUALIZATION WITH GGPLOT2



OUTLINE

1. Basics
2. Geometries
3. Aesthetics
4. Facets

LEARNING OBJECTIVE

Learn how to create multivariate visualizations with ggplot2.

R GRAPHICS

As of 2018, there are 4 graphical systems available in R:

- *base*
- *grid*
- *lattice*
- *ggplot2*

Access to the 4 systems differ: *base*, *grid*, *lattice* are included in the base installation; *grid*, *lattice*, *ggplot2* have to be loaded explicitly before being used.

A GGPLOT2 PRIMER

ggplot2 is a set of tools that map data to visual display elements, and that allow the user to control the fine details of 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

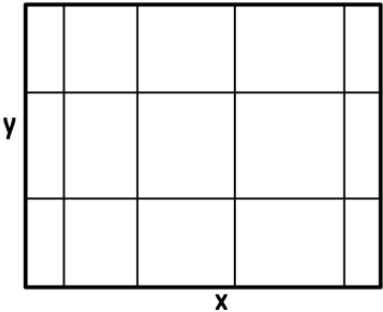
```
p <- ggplot(data = gapminder, mapping =  
  aes(x = gdp, y = lifexp, size = pop,  
      color = continent))
```

3. Geom

```
p + geom_point()
```

4. Co-Ordinates & Scales

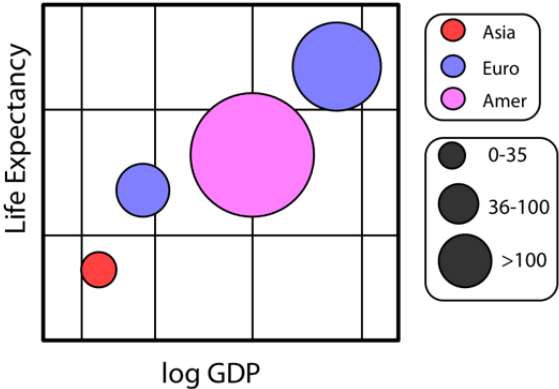
```
p + coord_cartesian() + scale_x_log10()
```



5. Labels & Guides

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

A Gapminder Plot



GGPLOT2 GRAMMAR – GEOMS

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 geoms.

GGPLOT2 GRAMMAR – GEOMS

Function	Adds	Options
<code>geom_bar()</code>	Bar chart	color, fill, alpha
<code>geom_boxplot()</code>	Box plot	color, fill, alpha, notch, width
<code>geom_density()</code>	Density plot	color, fill, alpha, linetype
<code>geom_histogram()</code>	Histogram	color, fill, alpha, linetype, binwidth
<code>geom_hline()</code>	Horizontal lines	color, alpha, linetype, size
<code>geom_jitter()</code>	Jittered points	color, size, alpha, shape
<code>geom_line()</code>	Line graph	color, alpha, linetype, size
<code>geom_point()</code>	Scatterplot	color, alpha, shape, size
<code>geom_rug()</code>	Rug plot	color, side
<code>geom_smooth()</code>	Fitted line	method, formula, color, fill, linetype, size
<code>geom_text()</code>	Text annotations	Many; see the help for this function
<code>geom_violin()</code>	Violin plot	color, fill, alpha, linetype
<code>geom_vline()</code>	Vertical lines	color, alpha, linetype, size

GGPLOT2 GRAMMAR – GEOMS

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()
```

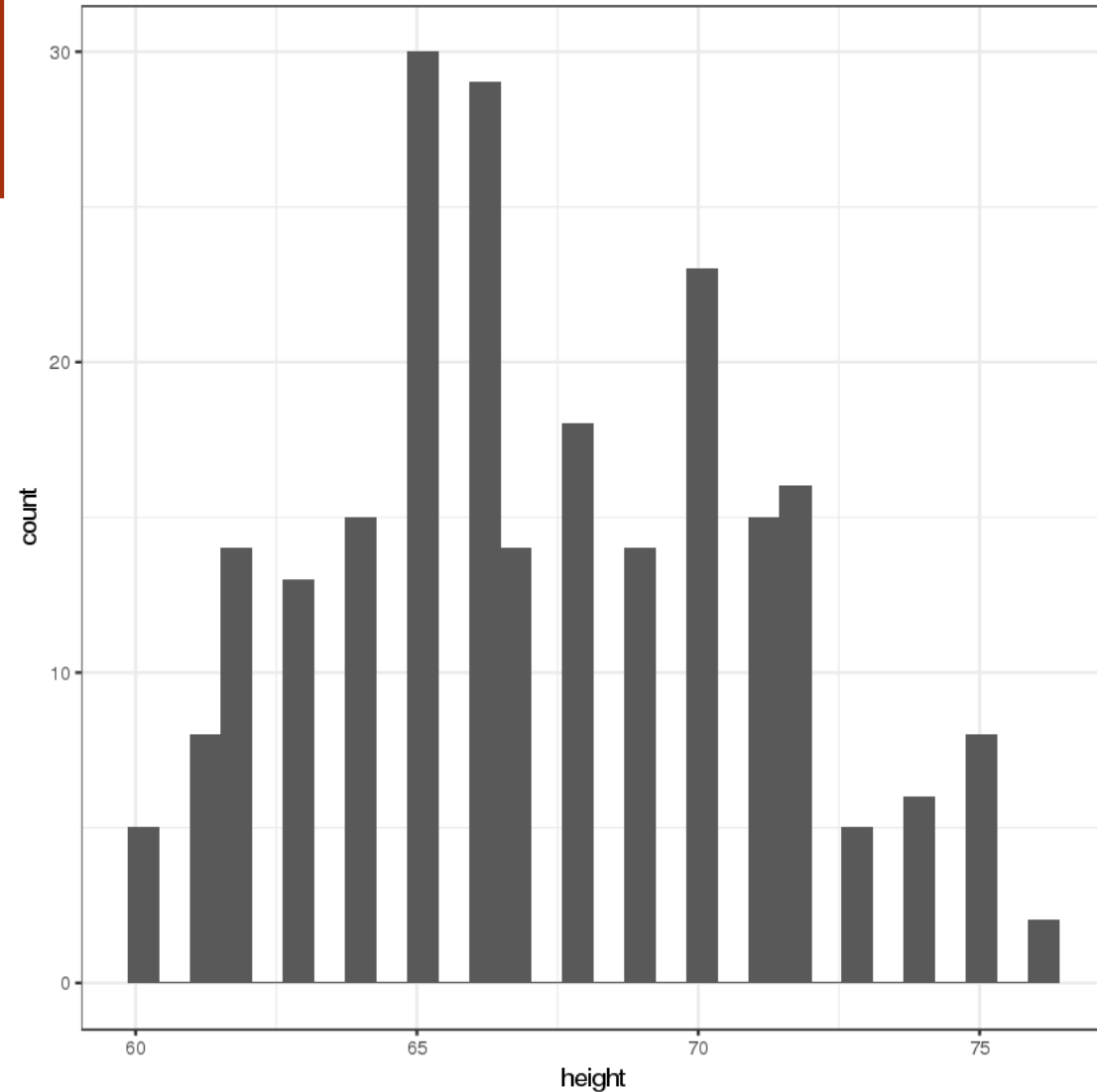
What do you expect the output to be?

GGPLOT2 GRAMMAR – 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()
```

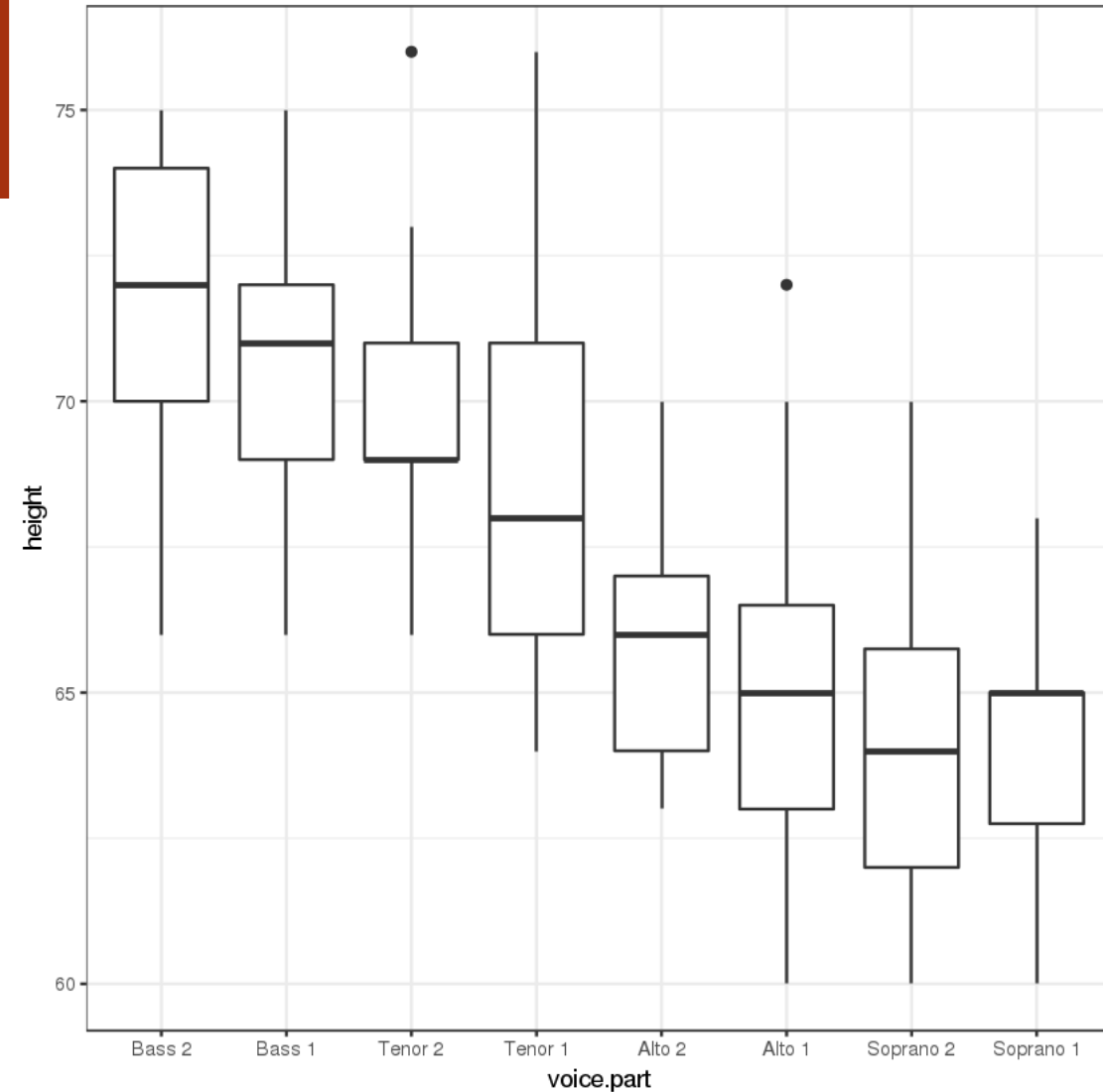


GGPLOT2 GRAMMAR – GEOM()

```
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()
```



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

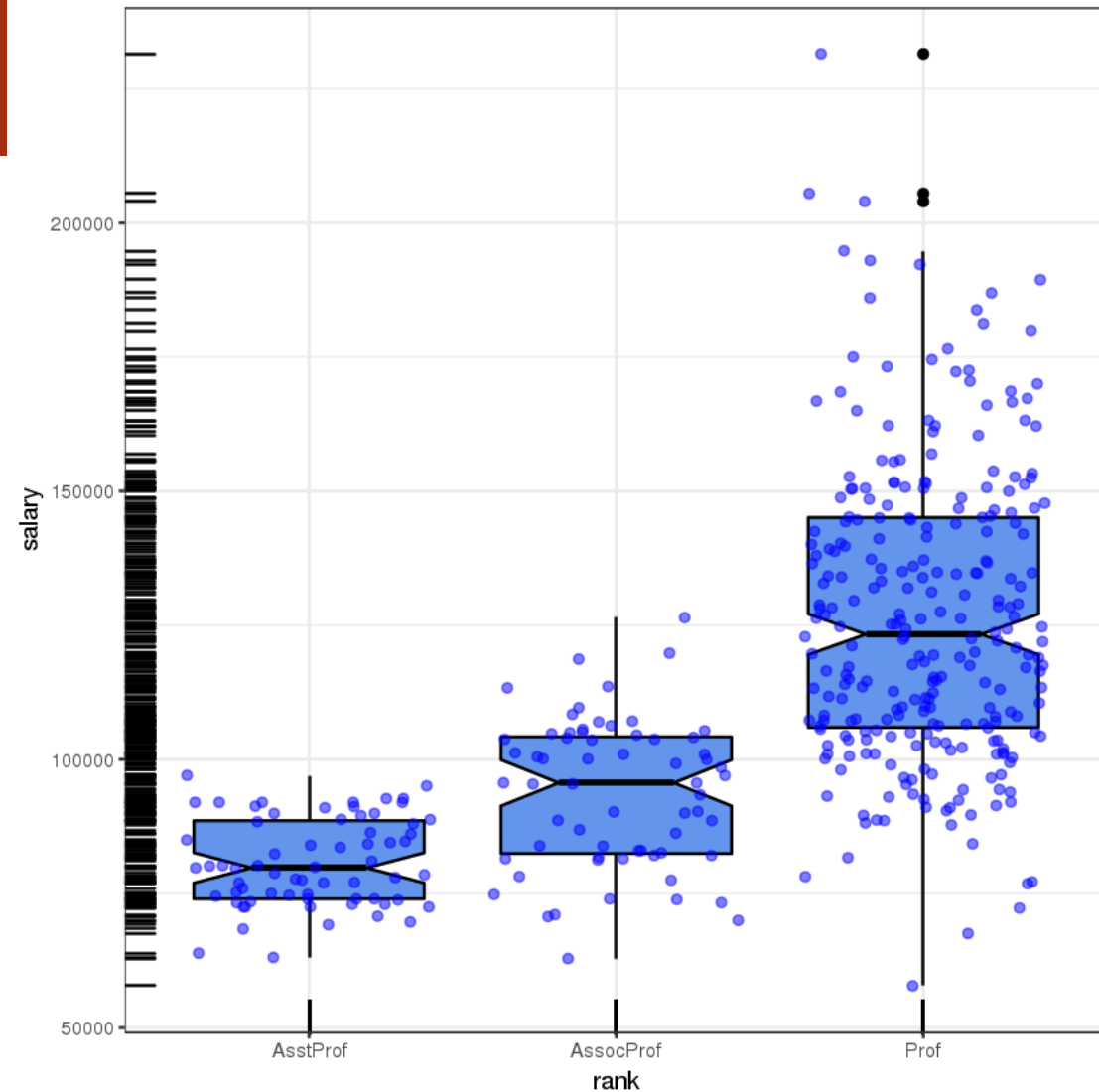
ggplot(Salaries, aes(x=rank, y=salary)) +
  geom_boxplot(fill="cornflowerblue", color="black", notch=TRUE) +
  geom_point(position="jitter", color="blue", alpha=.5) +
  geom_rug(side="l", color="black")
```

What do you expect the output to be?

GGPLOT2 GRAMMAR – 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")
```



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.

Aesthetics can be specified within the data function or within a geom. If they're specified within the data function then they apply to all specified geoms.

GGPLOT2 GRAMMAR – AESTHETICS

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

- `x, y, alpha, color, fill, shape, size`

Important difference between specifying characteristics (like colour and shape) inside and outside the `aes()` function

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

GGPLOT2 GRAMMAR – AES()

```
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")
```

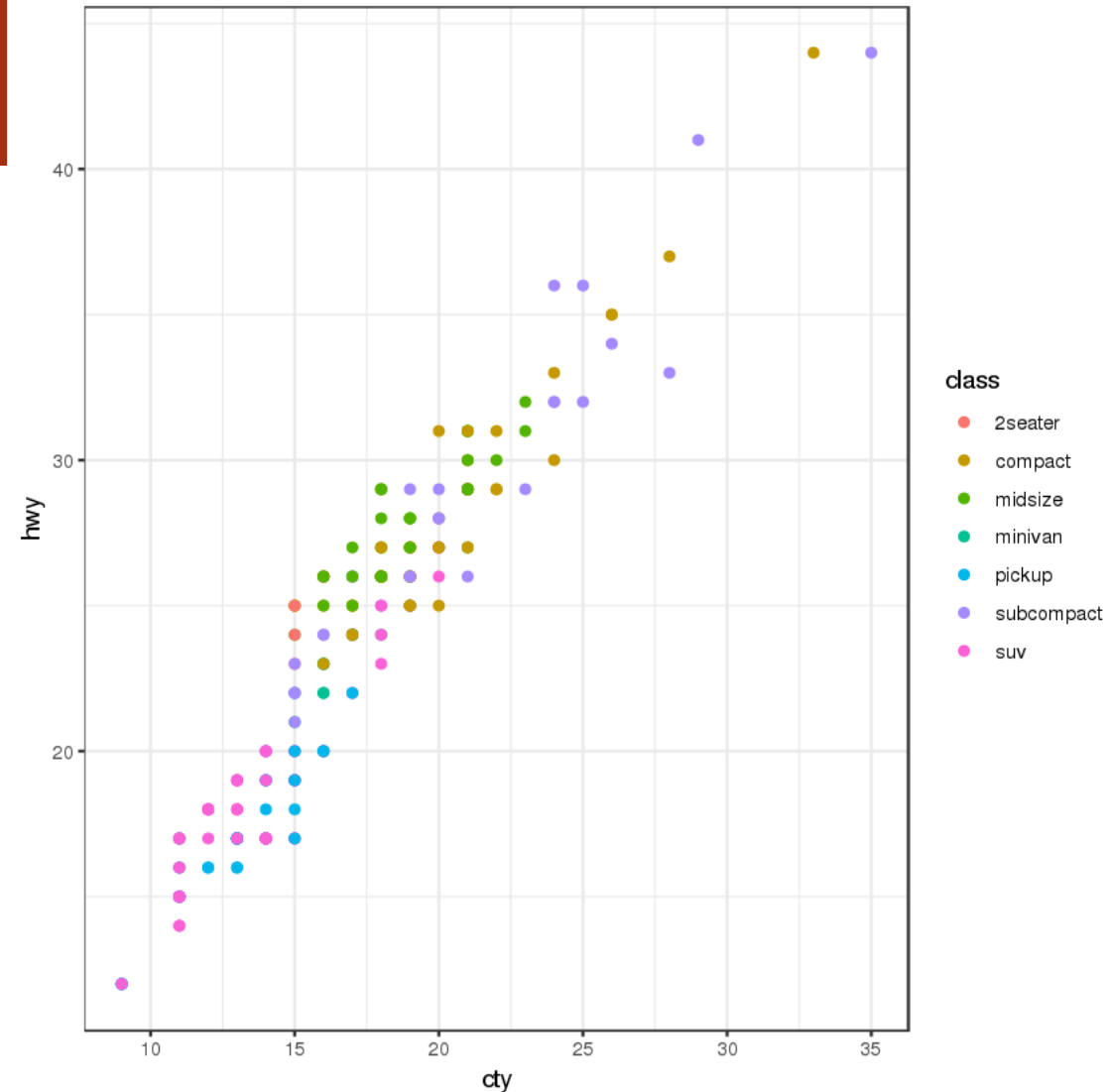
What do you expect the output to be?

GGPLOT2 GRAMMAR – AES()

```
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")
```

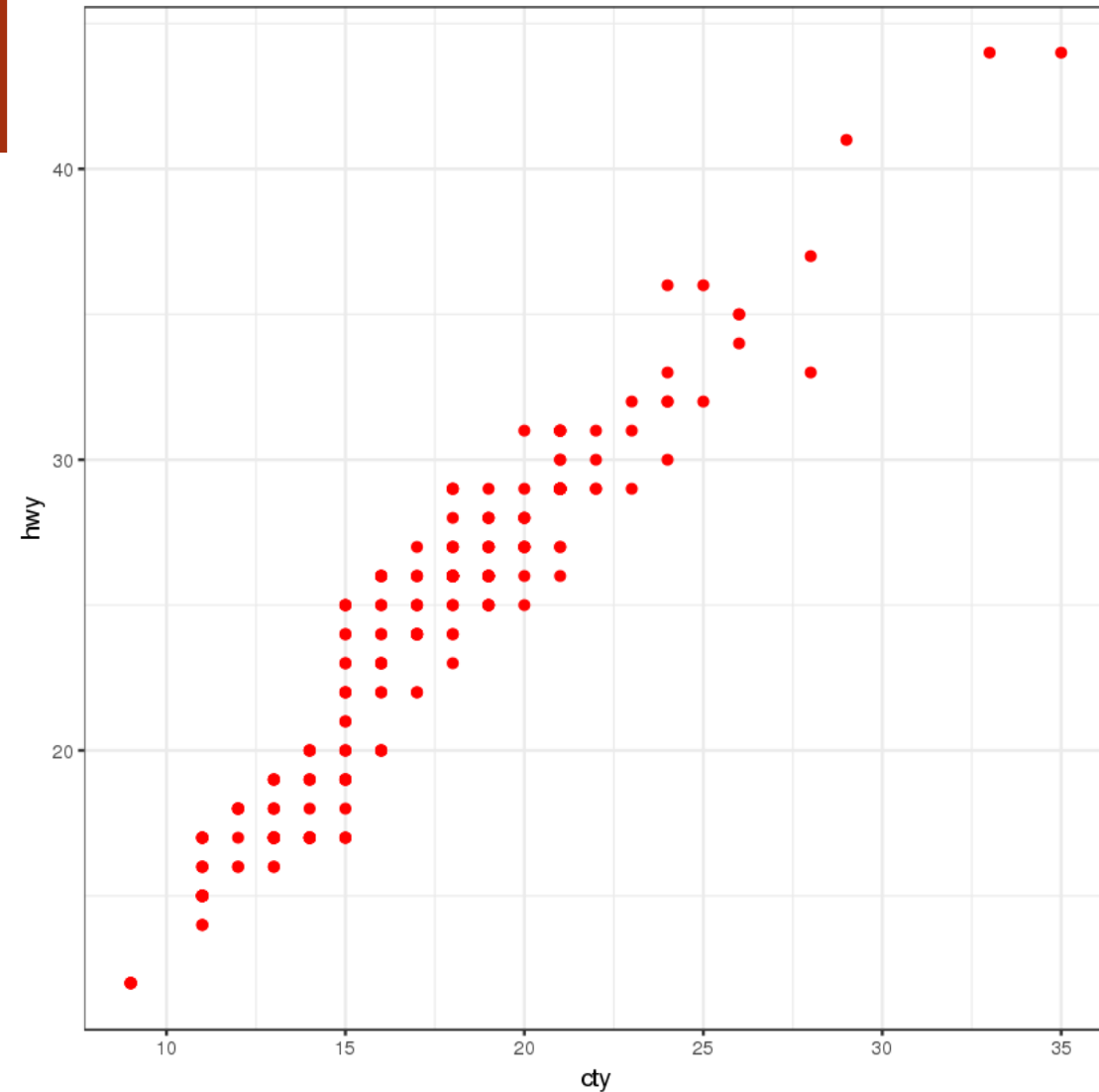


GGPLOT2 GRAMMAR – AES()

```
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.

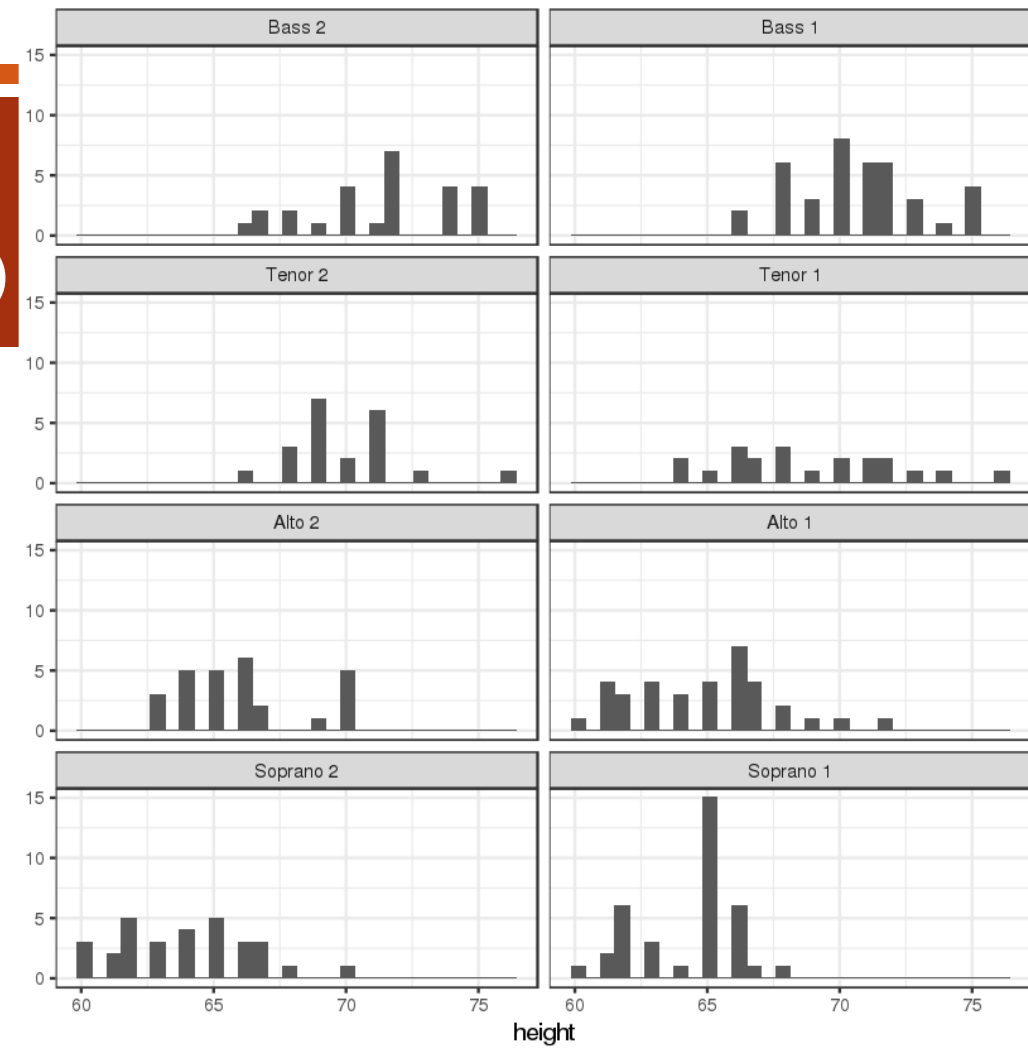
GGPLOT2 GRAMMAR – FACET_WRAP()

```
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 – FACET_WRAP()

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



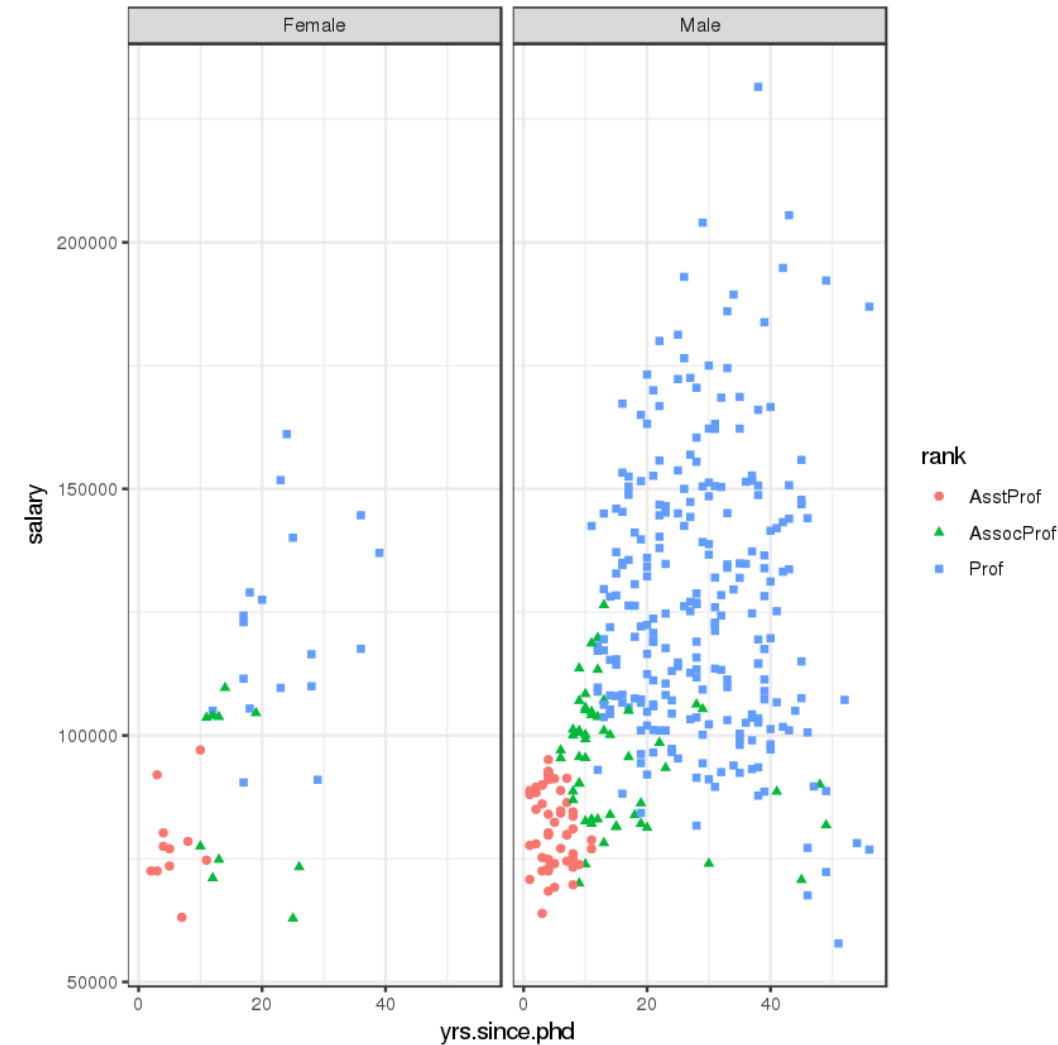
GGPLOT2 GRAMMAR – FACET_GRID()

```
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 – FACET_GRID()

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



EXERCISES

Detailed explanations and examples can be found in *A ggplot2 Primer*.

<https://www.data-action-lab.com/2018/11/12/a-ggplot2-primer/>

With a classmate, run through some of the examples provided in the R Notebook. It is not necessary for you to understand all the details of the visualizations, but it is useful to see what kind of outcomes are possible.

Select a dataset online, and generate 5 ggplot2 visualizations for it.

REFERENCES

REFERENCES

Tufte, E. [2001], *The Visual Display of Quantitative Information*, Graphics Press.

Hu, D. [1954], *How to Lie With Statistics*, Norton

Tufte, E. [2008], *Beautiful Evidence*, Graphics Press

Prabhakaran, S., [Top 50 ggplot2 Visualizations](#) (with Master List R Code).

Wickham, H. [2016], *ggplot2: Elegant Graphics for Data Analysis* (2nd ed), Springer.

Chang, W. [2013], *R Graphics Cookbook*, O'Reilly.

Wickham, H. [2009], A Layered Grammar of Graphics, *Journal of Computational and Graphical Statistics* 19:3–28.

REFERENCES

Nussbaumer Knaflic, C. [2015], *Storytelling with Data*, Wiley

Cairo, A. [2013], *The Functional Art*, New Riders

Cairo, A. [2016], *The Truthful Art*, New Riders

Meireilles, I. [2013], *Design for Information*, Rockport

50 Great Examples of Data Visualization: <http://www.webdesignerdepot.com>

Nathan Yau's [FlowingData](#)

[Data Visualization](#) on Wikipedia

[Misleading Graphs](#) on Wikipedia

REFERENCES

Kabacoff, R.I. [2011], R in Action, Second Edition: Data analysis and graphics with R, Live.

Maindonald, J.H. [2008], Using R for Data Analysis and Graphics: Introduction, Code and Commentary.

Tyner, S., Briatte, F., Hofmann, H. [2017], Network Visualization with ggplot2, The R Journal, vol. 9(1).

Broman, K. [2016], Data Visualization with ggplot2.

Robinson, D., Visualizing Data Using ggplot2, on varianceexplained.org.

Manipulating, analyzing and exporting data with tidyverse, on datacarpentry.org.

Wickham, H. [2014], Tidy Data, Journal of Statistical Software, v59, n10.

Gashim, E., Boily, P. [2018], A ggplot2 Primer, data-action-lab.com