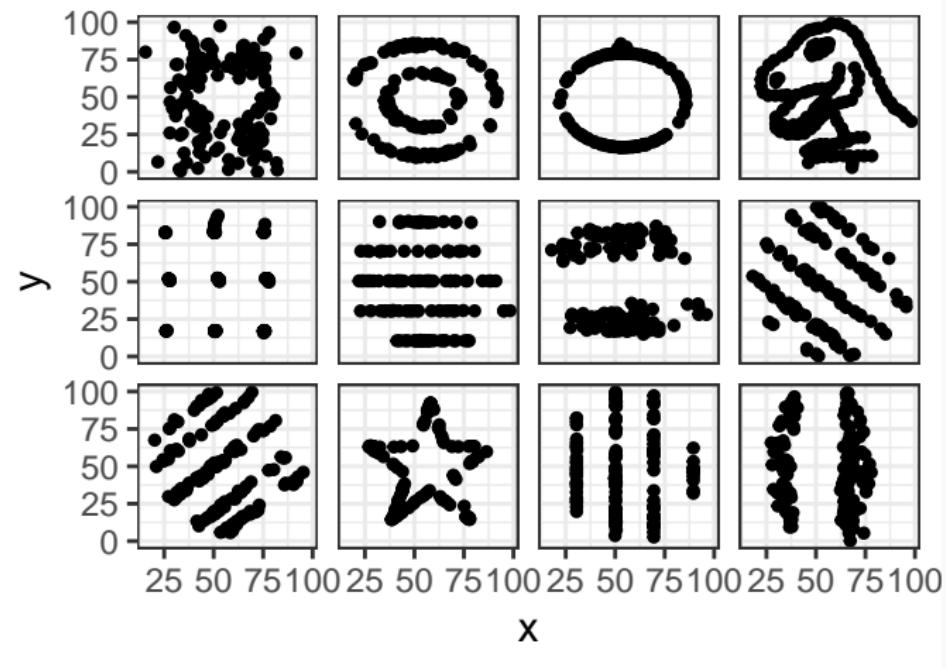


Data visualisation with ggplot2

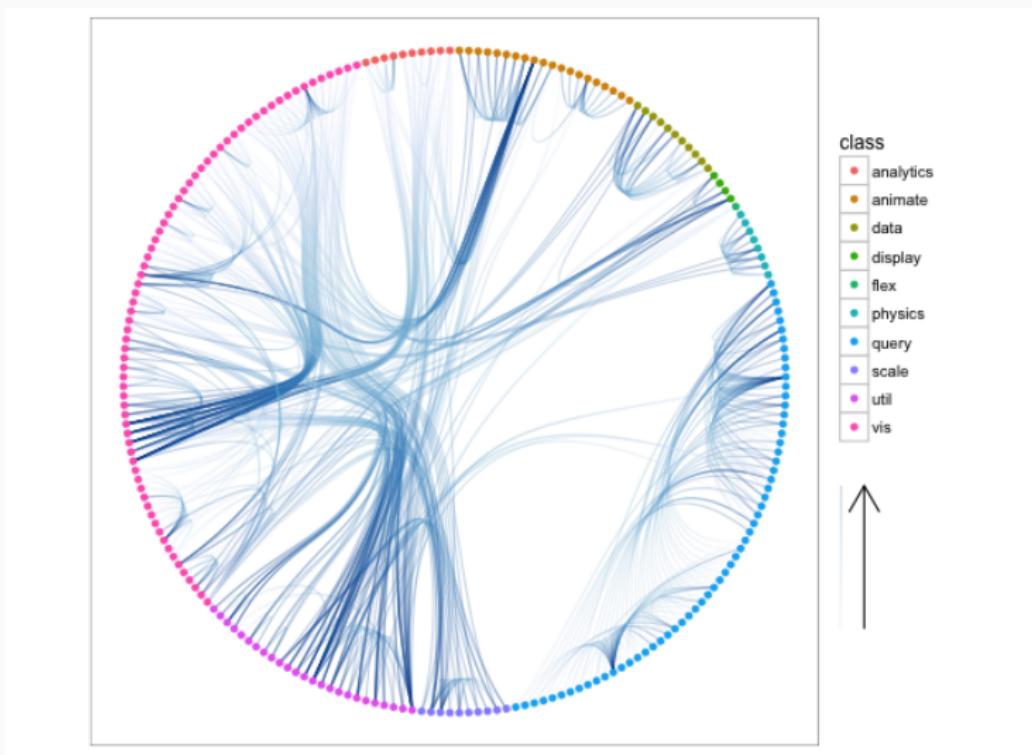
Francisco Rodriguez-Sanchez (@frod_san)

Always plot data!



<https://github.com/stephlocke/datasauRus>

Made with ggplot



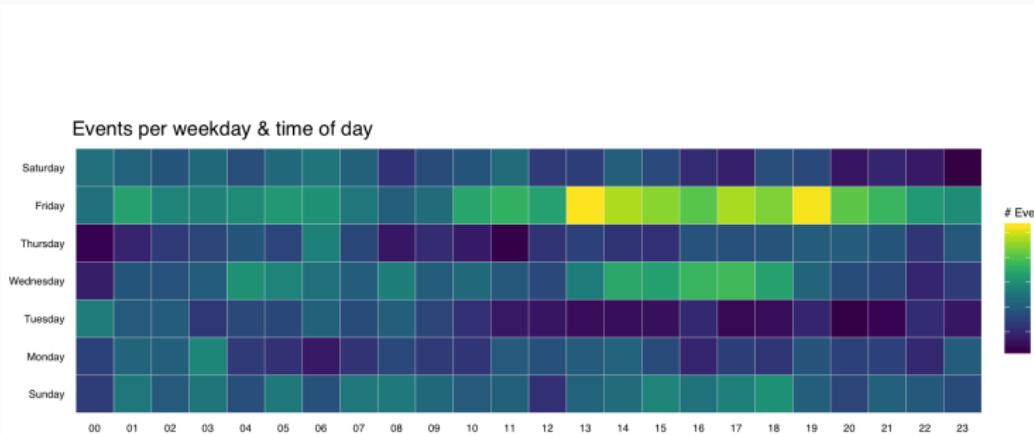
<https://github.com/thomasp85/ggraph>

Made with ggplot



<http://spatial.ly/2012/02/great-maps-ggplot2/>

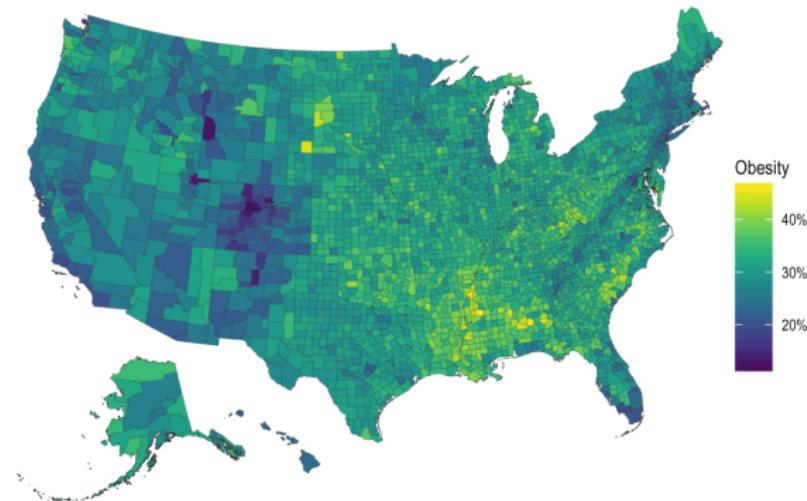
Made with ggplot



<https://rud.is/b/2016/02/14/making-faceted-heatmaps-with-ggplot2/>

U.S. Obesity Rate by County (2012)

Content source: Centers for Disease Control and Prevention

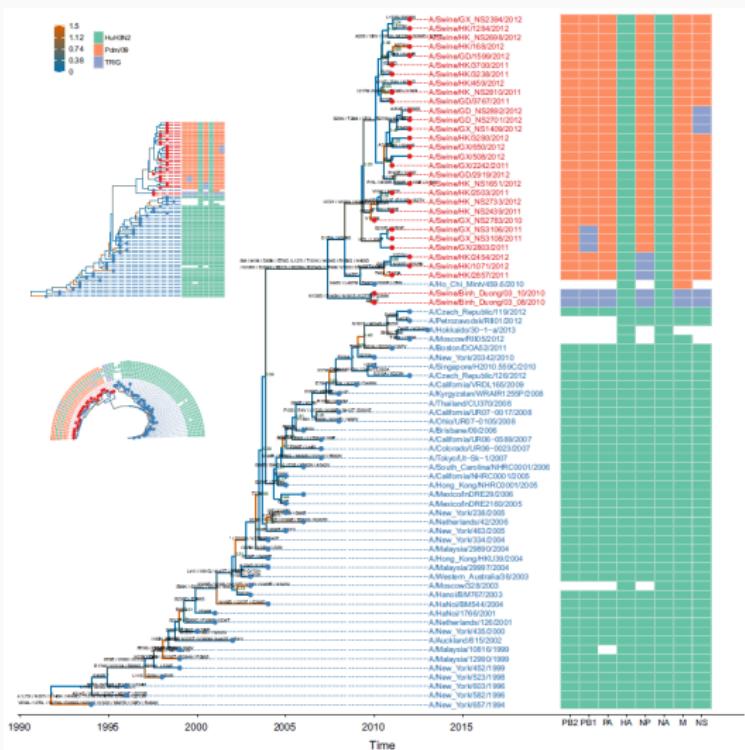


Data from http://www.cdc.gov/diabetes/atlas/countydata/County_ListofIndicators.html

[https:](https://rud.is/b/2016/03/29/easier-composite-u-s-choropleths-with-albersusa/)

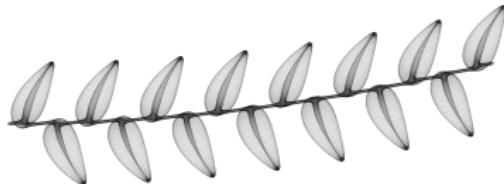
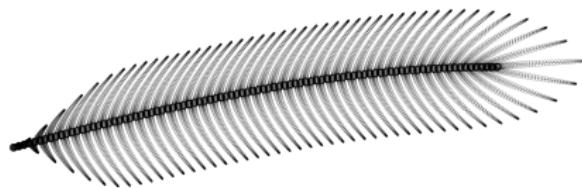
//rud.is/b/2016/03/29/easier-composite-u-s-choropleths-with-albersusa/

Made with ggplot



<https://guangchuangyu.github.io/ggtree/>

Made with ggplot



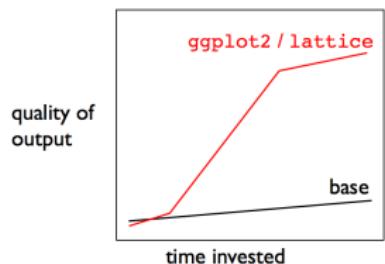
<https://github.com/marcusvolz/mathart>

Why ggplot

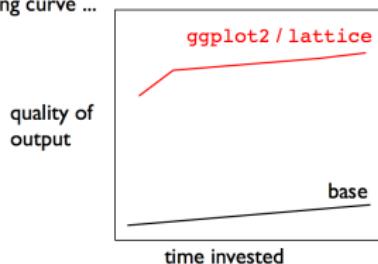
- Extremely powerful and flexible
- Consistent (grammar of graphics)
- Very powerful user base and active development

At the beginning it's hard, but then it pays off

week one



after you've climbed the steepest part of the learning curve ...



* figure is totally fabricated but, I claim, still true

* figure is totally fabricated but, I claim, still true

Source: <https://github.com/jennybc/ggplot2-tutorial>

Very good documentation and tutorials

- Official `ggplot2` documentation
- `ggplot2` book
- `R graphics cookbook` and `Cookbook for R`
- Beautiful plotting in R: A `ggplot2` cheatsheet
- Introduction to `ggplot2`
- Tutorial: `ggplot2`
- How to format plots for publication using `ggplot2`
- Visualising data with `ggplot2`
- Data Visualization with R and `ggplot2`
- `ggplot2` tutorial
- Data visualisation chapter in R for Data Science
- The complete `ggplot2` tutorial
- Data visualization: a practical introduction (K. Healy)
- Fundamentals of data visualization (C. Wilke)

Cheatsheet

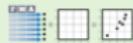
Data Visualization with ggplot2

Cheat Sheet



Basics

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



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



Complete the template below to build a graph.

```
ggplot([data = mpg], aes(x = cyl, y = hwy))  
  + [ ]  
  + stat = [ ]  
  + position = [ ]  
  + [ ]  
  + [ ]  
  + [ ]  
  + [ ]  
  + [ ]
```

Required
Not required, sensible defaults supplied

Begin a plot that you finish by adding layers to. Add one geom function per layer.

```
ggplot([x = cyl, 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.  
ggplot("plot.png", width = 5, height = 5)  
  Saves last plot as 5" x 5" file named "plot.png" in working directory. Makes 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 + `geom_abline(aes(xintercept = 0), intercept = 1)`
- b + `geom_blank()` (useful for expanding limits)
- c + `geom_curve(aes(cx = lat + 1, cxend = lat - 1, cy = check, cyend = overlap), check = TRUE)`
- d + `geom_jitter(aes(x = 2, width = 2))`
- e + `geom_point(aes(x = 2, fill = shape, size = 2))`
- f + `geom_quartz()`
- g + `geom_rug(aes(tick = 10))`
- h + `geom_smooth(aes(method = "loess"))`
- i + `geom_ribbon(aes(ymin = unempoly - 900, ymax = unempoly + 900))`

Line Segments

- common aesthetics: x, y, alpha, color, linetype, size
- b + `geom_abline(aes(intercept = 0), slope = 1)`
 - c + `geom_hline(aes(intercept = 0))`
 - d + `geom_vline(aes(intercept = 0))`
 - e + `geom_segment(aes(x0 = 1, xend = 2, y0 = 3, yend = 4))`
 - f + `geom_spline(aes(closed = TRUE, radius = 1))`

One Variable

- Continuous
- c + `geom_area(aes(fill = "white"))`
 - c + `geom_area(stat = "bin")`
 - c + `geom_density(aes(fill = "gaussian"))`
 - c + `geom_dotplot()`
 - c + `geom_freqpoly()`
 - c + `geom_histogram(aes(binwidth = 5))`
 - c + `geom_qq(aes(sample = hwy))`
 - c + `geom_violin(aes(fill = "white"))`

Discrete

- c + `geom_bar()`
- d + `geom_pie(aes(fill = "white"))`

Two Variables

Continuous X, Continuous Y

- e + `geom_label(aes(label = "hi", nudge_x = 1, nudge_y = 1, check_overlap = TRUE))`
- f + `geom_rect(aes(x = 1, xend = 2, y = 1, yend = 2))`
- g + `geom_text(aes(x = 1, y = 1))`

Continuous Function

- i + `geom_area()`
- i + `geom_line()`
- i + `geom_step(direction = "hv")`

Continuous Bivariate Distribution

- b + `geom_bivar2d(aes(meshwidth = 0.025, ncells = 500))`
- b + `geom_hex()`

Visualizing error

- d + `data frame(p1 = rnorm(1000, 0, 1), p2 = rnorm(1000, 0, 2))`
- j + `ggplot(d1, aes(p1, p2, fill = abs(p1 - p2)))`

Discrete X, Continuous Y

- f + `geom_col()`
- f + `geom_boxplot()`
- x, y, lower, middle, upper, ymin, ymax, alpha, color, fill, group, linetype, shape, weight
- g + `geom_dots(aes(y = "y"))`
- g + `geom_pointrange(aes(y = "y"))`
- g + `geom_violin(aes(fill = "white"))`

Discrete X, Discrete Y

- c + `geom_raster()`
- c + `geom_count()`
- c + `geom_bar()`

Maps

- data + `data frame(state = c("USArrests$Murder, state = USArrests$Murder))`
- map + `map_data("state", "us"))`
- k + `geom_map(aes(map_id = state, map = map))`
- k + `expand_limits(x = mapling, y = maplat)`

Three Variables

- sealidz + `sealidz, opacity_dz, alpha_dz, longitude + latitude, id, height = 0.5,`
1 + `ggplot(sealidz, aes(id, longitude, latitude))`
- l + `geom_contour(aes(z = 2))`
- l + `geom_tile(aes(z = z))`

Learn more at ggplot2.org and ggplot2-exts.org • ggplot2 2.1.0 • Updated 11/16

<https://www.rstudio.com/resources/cheatsheets/>

Repos of figures + code

- R graph catalog
- From Data to Viz
- The R graph gallery
- R graph gallery
- Cookbook for R: Graphs
- Graphical data analysis with R
- IEG figures

Find answers for all your questions in Stack Overflow



Search

ggplot2

36,854 results



The Practical Dev
@ThePracticalDev



The last programming book you'll ever need

Cutting corners to meet arbitrary management deadlines



Essential

Copying and Pasting
from Stack Overflow

Building a ggplot figure

Our example dataset: tree heights and DBH

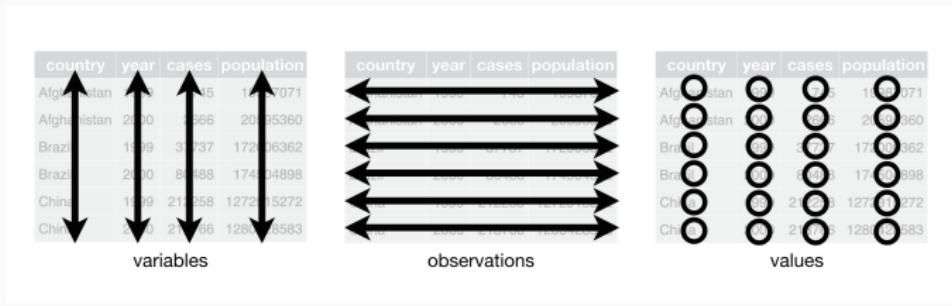
<http://tinyurl.com/treesdata>

- One species
- 10 plots
- 1000 trees
- Number of trees per plot ranging from 4 to 392

```
trees <- read.csv("data/trees.csv")
summary(trees[, 1:3])
```

	plot	dbh	height
Min.	: 1.0	Min. : 5.06	Min. :13.40
1st Qu.	: 1.0	1st Qu.:17.69	1st Qu.:29.68
Median	: 2.0	Median :28.62	Median :36.55
Mean	: 2.7	Mean :27.88	Mean :36.51
3rd Qu.	: 4.0	3rd Qu.:38.97	3rd Qu.:43.33
Max.	:10.0	Max. :49.92	Max. :59.30

Data must be a tidy data frame



```
tidy::gather(table4, key = "year", value = "cases", "1999", "2000")
```

The diagram shows the transformation of a wide table into a tidy format using the `tidy::gather` function.

The original wide table (`table4`) has columns for country, year, and cases. The cases column contains two values for each country: 745 for 1999 and 2666 for 2000. The resulting tidy table has columns for country, 1999, and 2000. The 1999 column contains the value 745 for Afghanistan and 37737 for Brazil. The 2000 column contains the value 2666 for Afghanistan and 80488 for Brazil. The country column identifies the rows for each country.

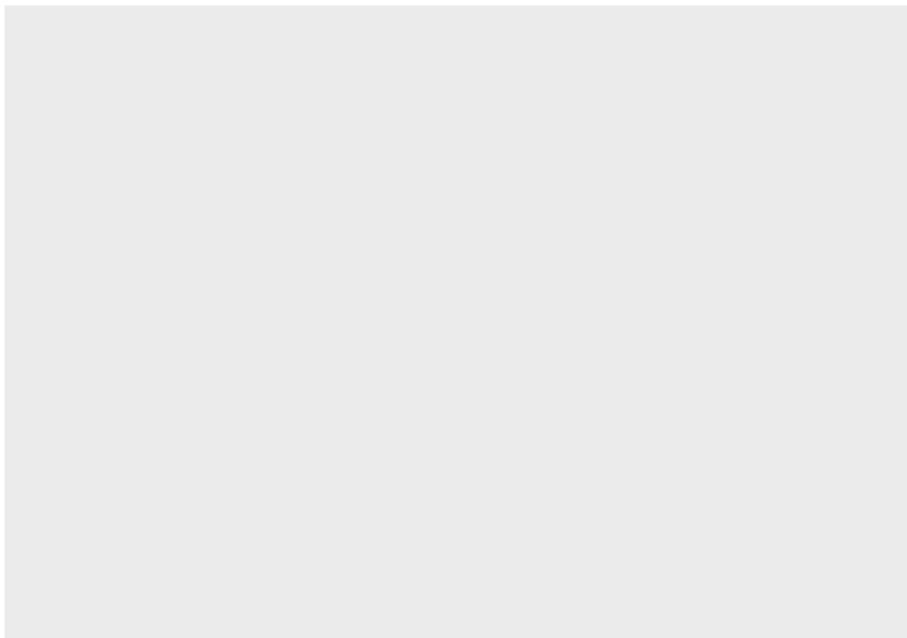
country	year	cases
Afghanistan	1999	745
Afghanistan	2000	2666
Brazil	1999	37737
Brazil	2000	80488
China	1999	212258
China	2000	213766

country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

<http://r4ds.had.co.nz/tidy-data.html>

Calling ggplot

```
library(ggplot2)  
ggplot(trees)
```

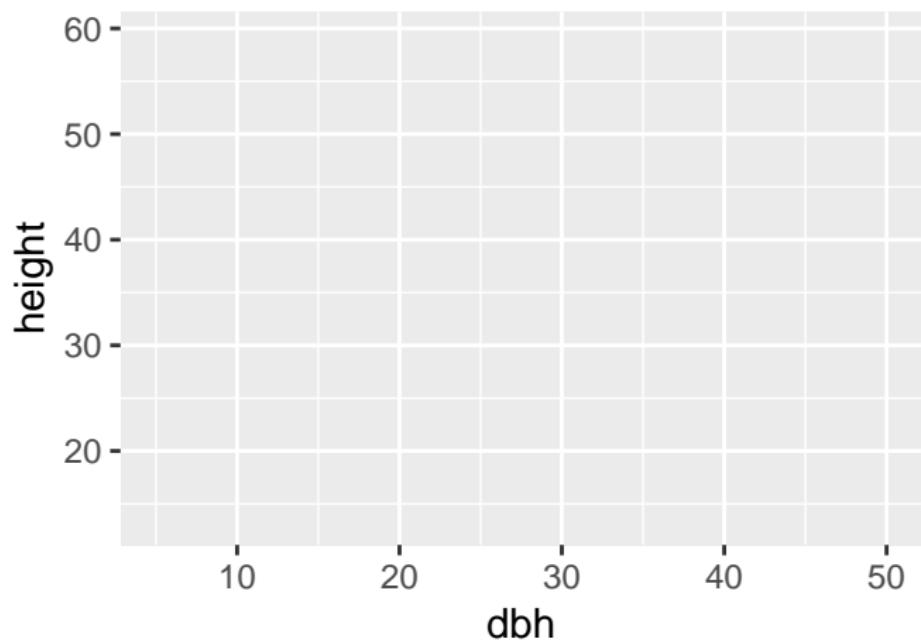


```
ggplot(trees)
```

First argument is a tidy data frame

What variables as axes?

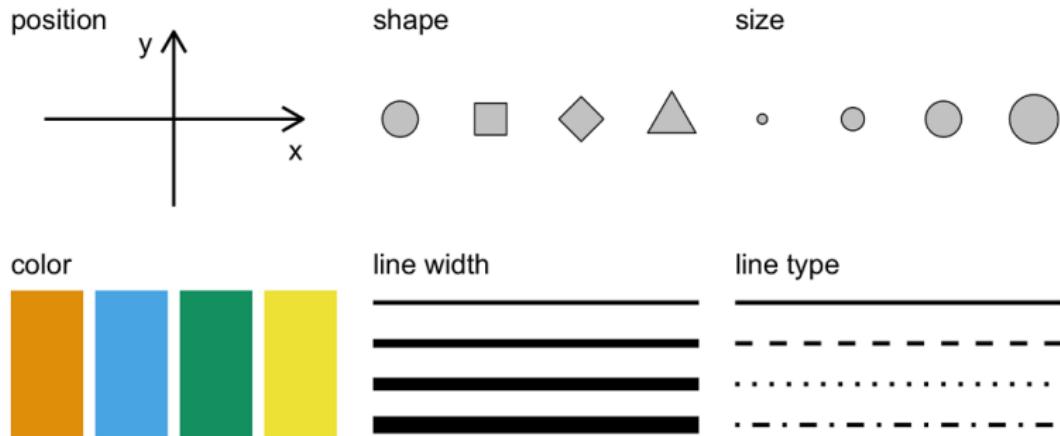
```
ggplot(trees) +  
  aes(x = dbh, y = height)
```



Note syntax: + followed by new line

```
ggplot(trees) +  
  aes(x = dbh, y = height)
```

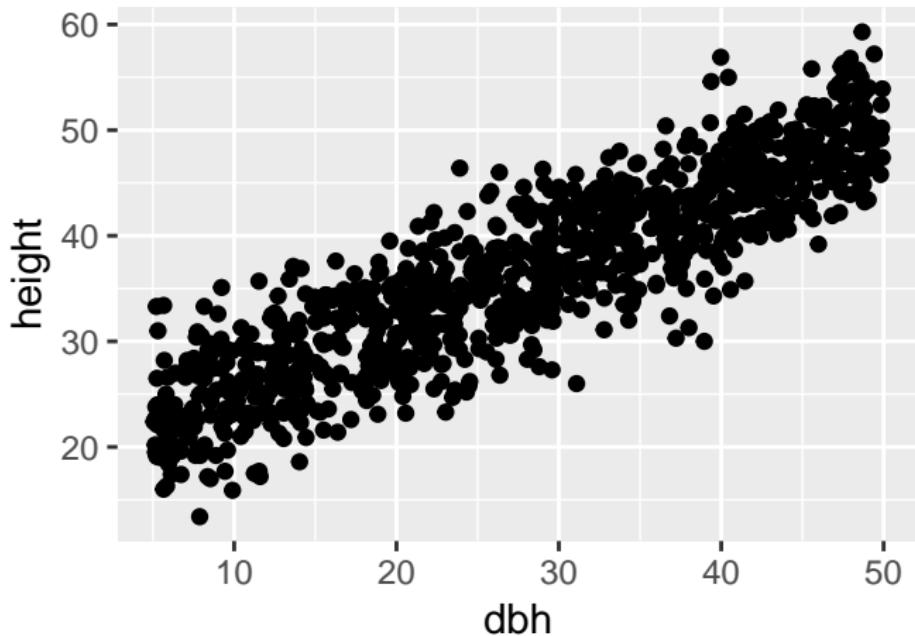
Aesthetics (`aes`) map data variables (`dbh`, `height`) to graphic elements (`axes`)



<http://serialmentor.com/dataviz/aesthetic-mapping.html>

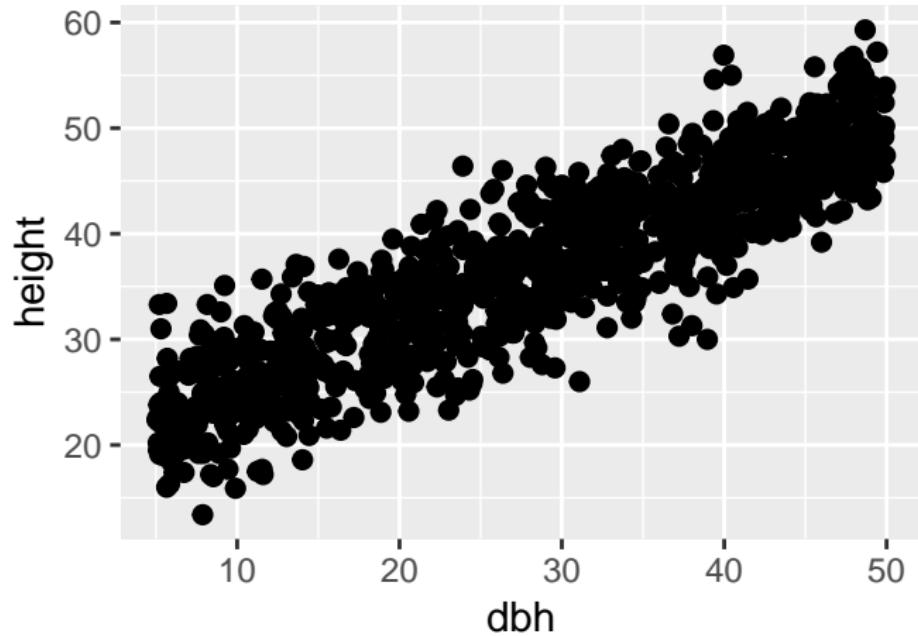
Adding layers (geoms)

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point()
```



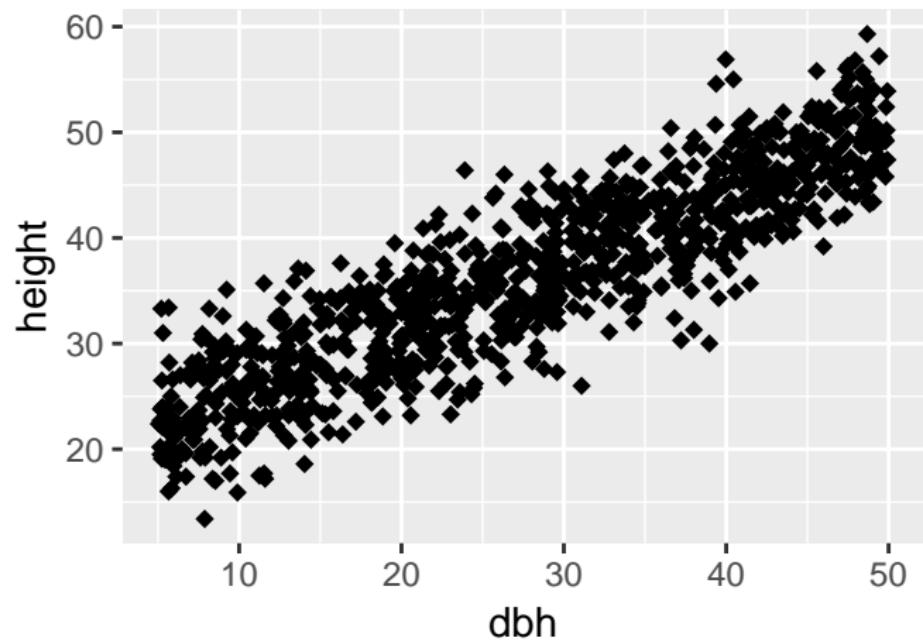
Changing point size

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(size = 2)
```



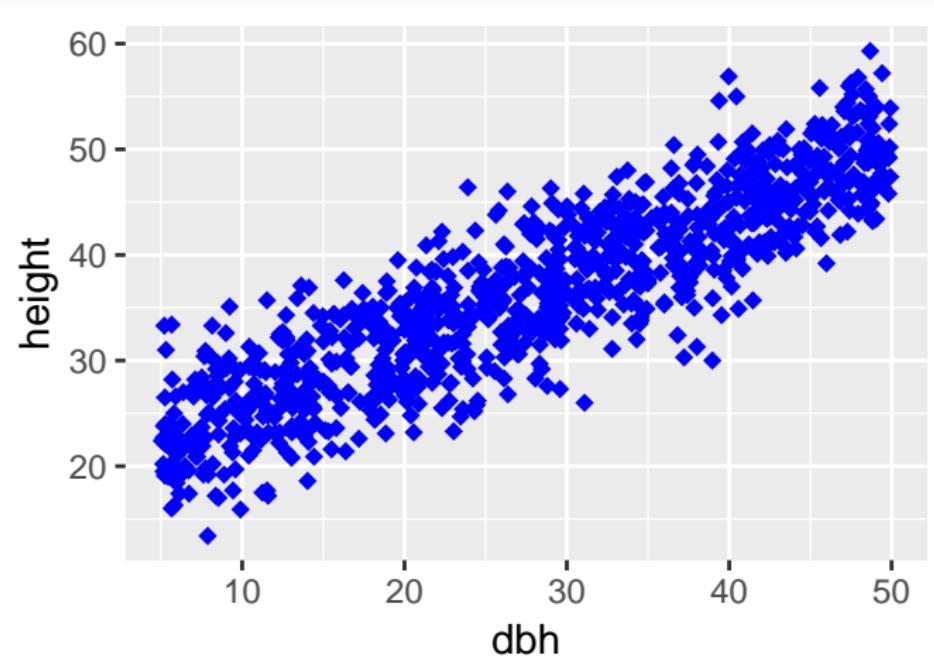
Changing point size and type

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(size = 2, shape = 18)
```



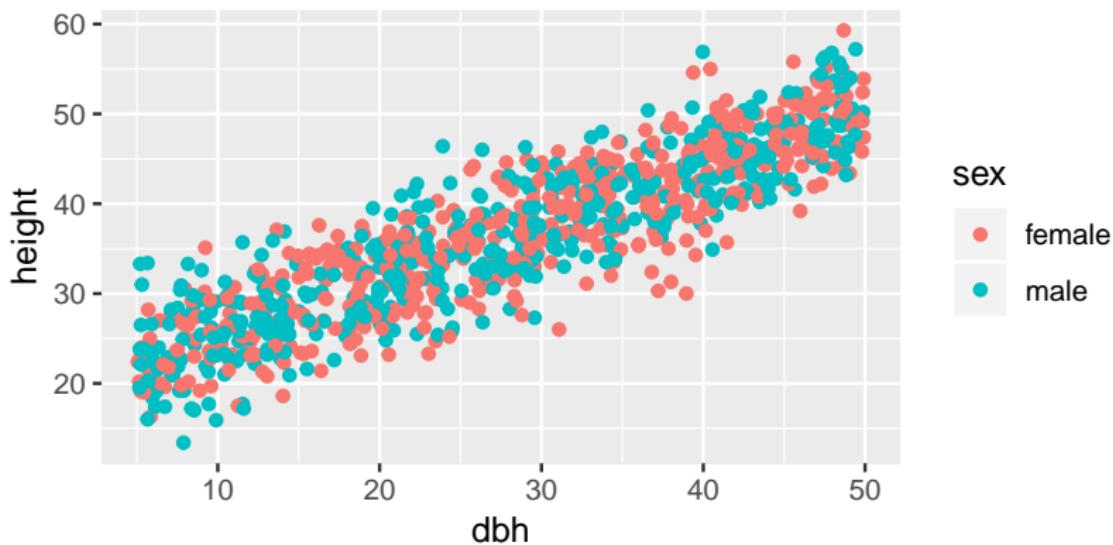
Changing point size and type

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(size = 2, shape = 18, colour = 'blue')
```



Map geom aesthetics (e.g. colour) to variable

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = sex))
```



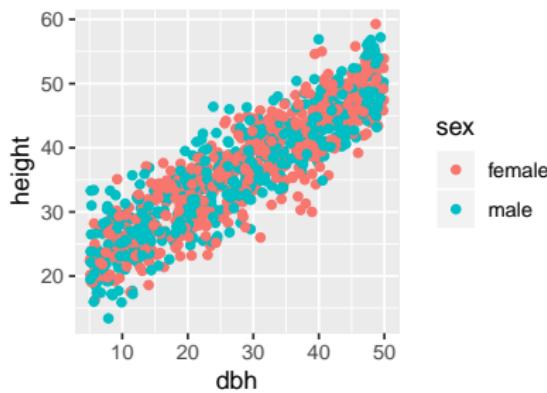
Note difference between

```
geom_point(colour = "blue")
# colour is given a concrete value ('blue')
```

```
geom_point(aes(colour = sex))
# colour maps a *variable* (using `aes`)
```

This works:

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = sex))
```



This doesn't work:

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(colour = sex)
```

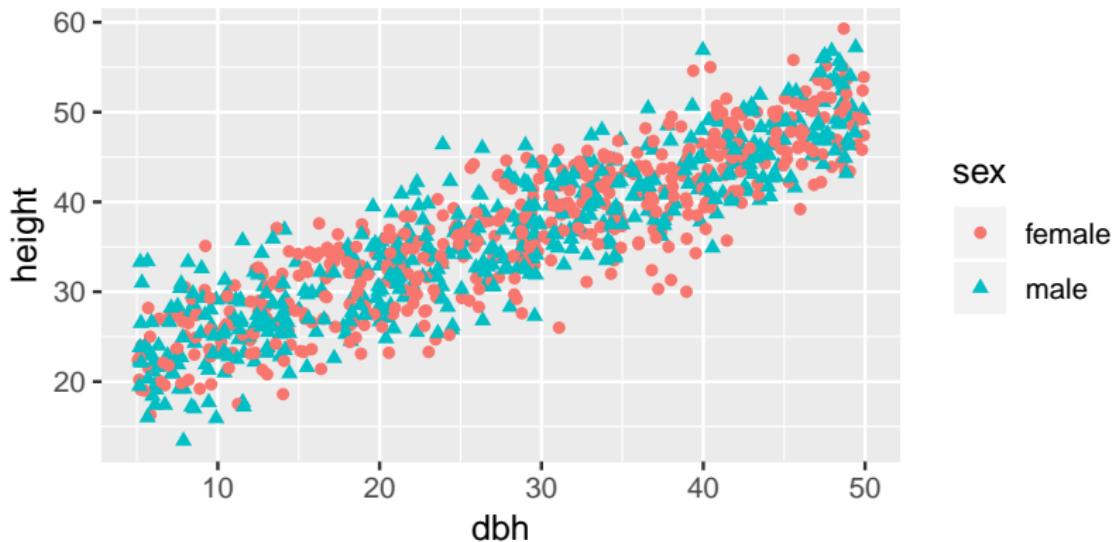
*Error in layer(data = data, mapping =
mapping, stat = stat, geom =
GeomPoint, : object 'sex' not found*

'sex' is a variable in dataframe

Must use aes

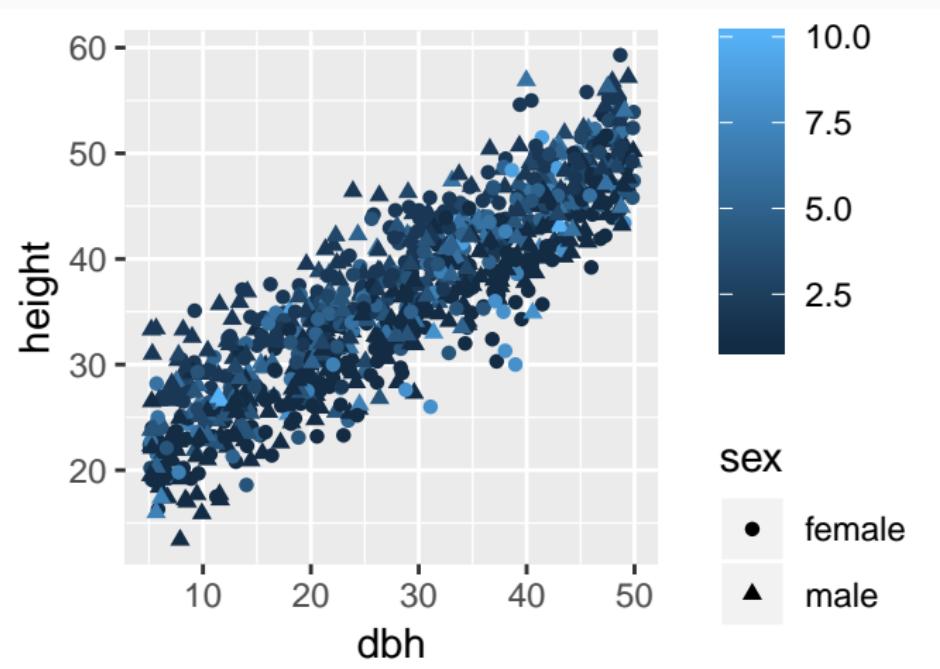
Map geom aesthetics (colour, shape) to variable

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = sex, shape = sex))
```



Map geom aesthetics (colour, shape) to variable

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = plot, shape = sex))
```

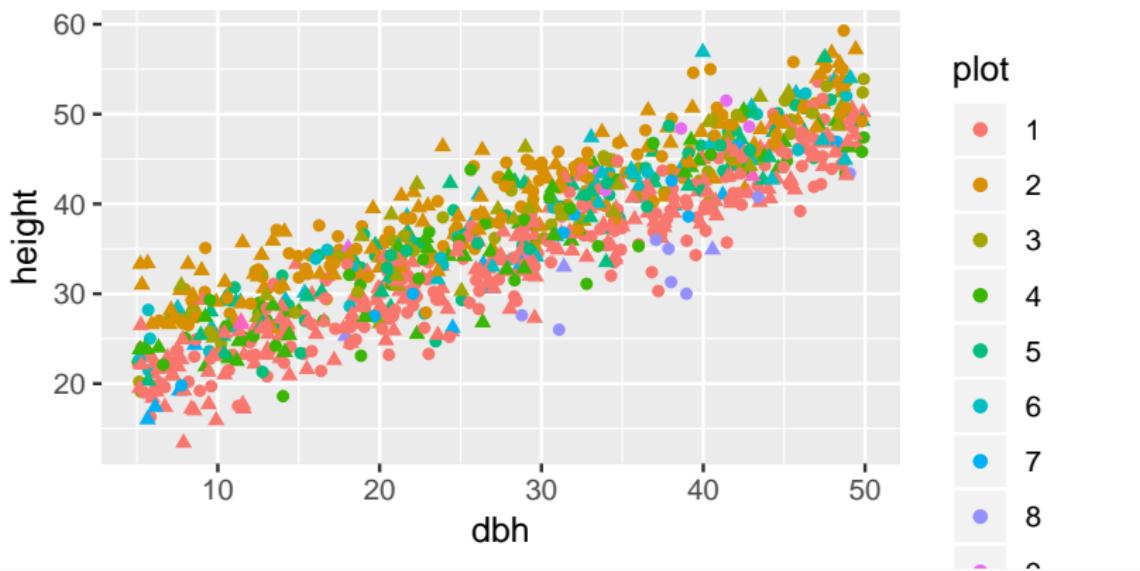


Ensuring 'plot' is a factor, not numeric

```
trees$plot <- as.factor(trees$plot)
```

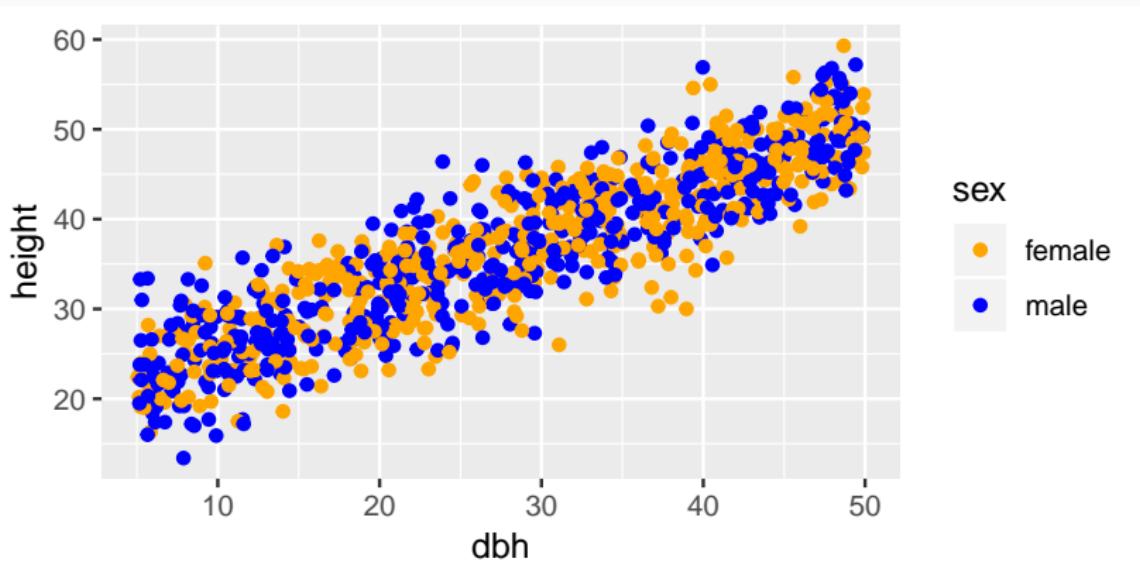
Map geom aesthetics (colour, shape) to variable

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = plot, shape = sex))
```



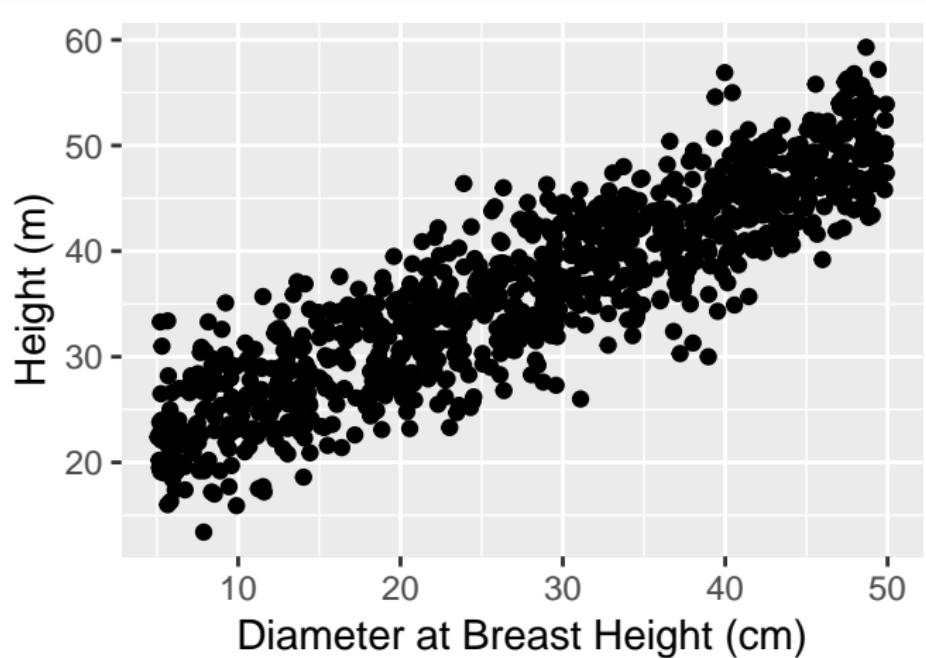
Change colour scale

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = sex)) +  
  scale_colour_manual(values = c("orange", "blue"))
```



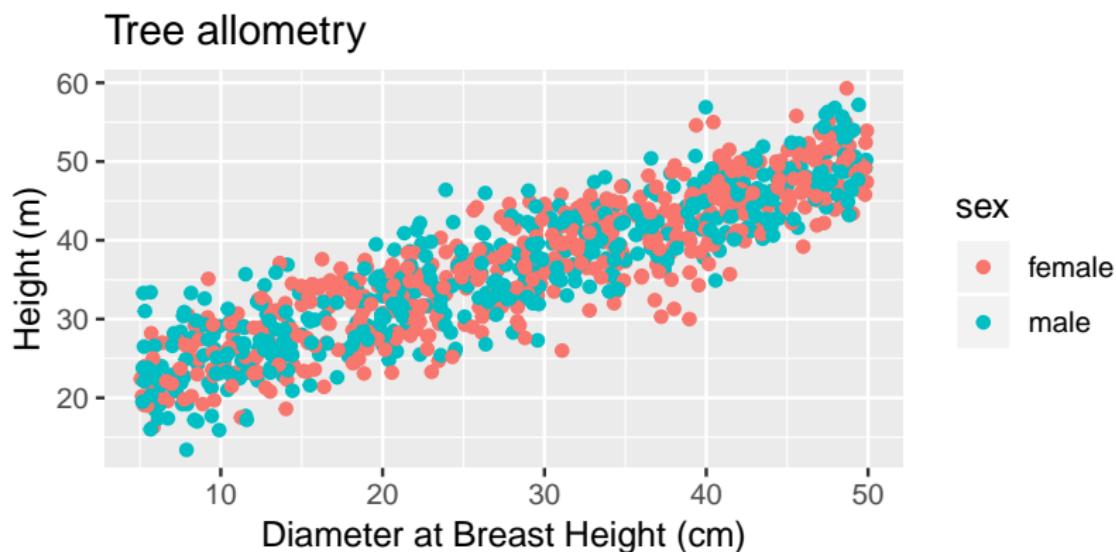
Change axis labels: `xlab` & `ylab`

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point() +  
  labs(x = "Diameter at Breast Height (cm)", y = "Height (m)")
```



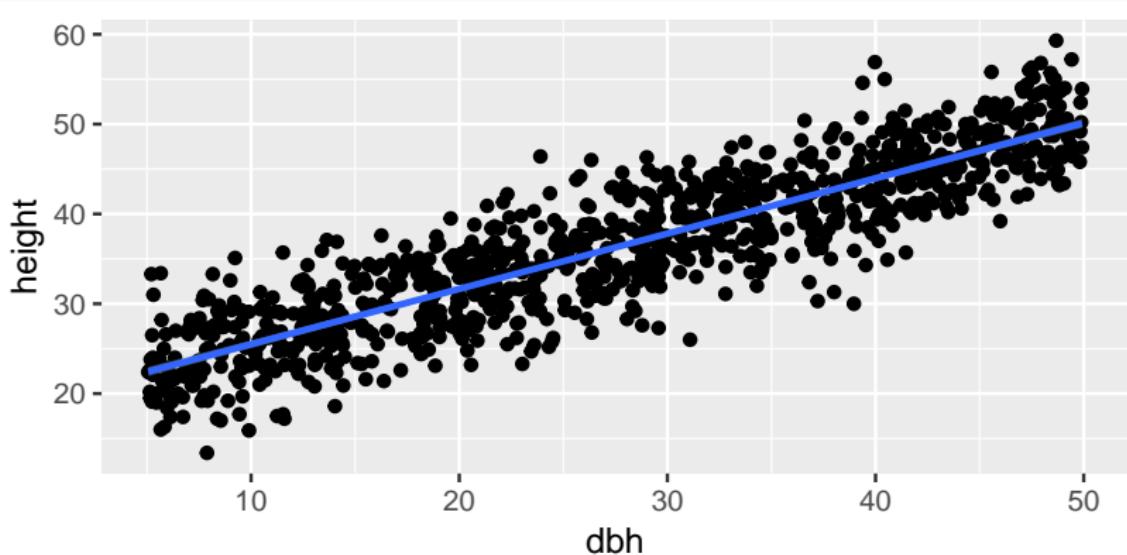
Set title

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point(aes(colour = sex)) +  
  labs(x = "Diameter at Breast Height (cm)", y = "Height (m)") +  
  labs(title = "Tree allometry")
```



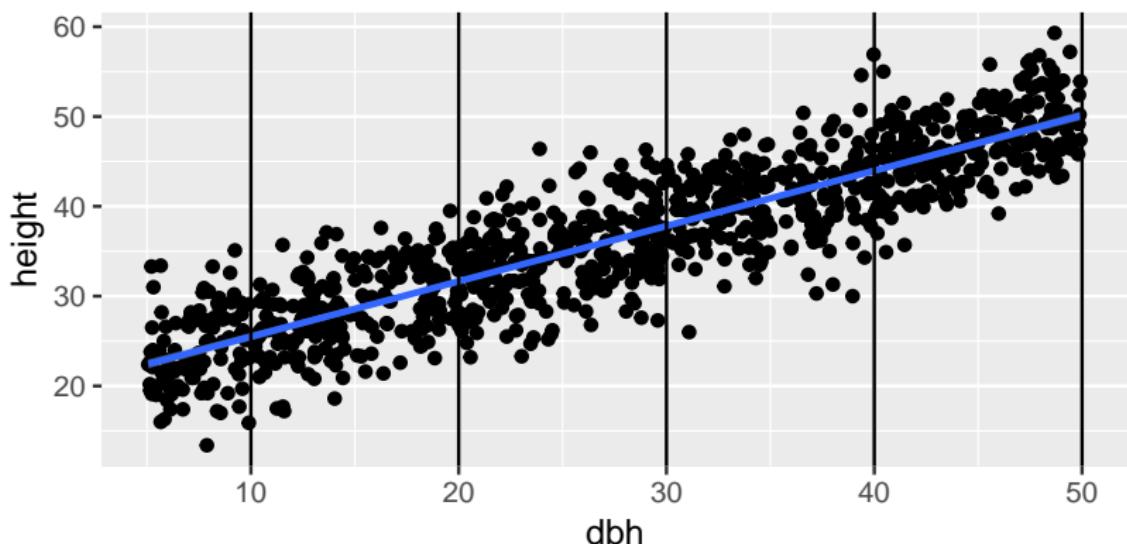
Adding more layers

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point() +  
  geom_smooth(method = "lm")
```



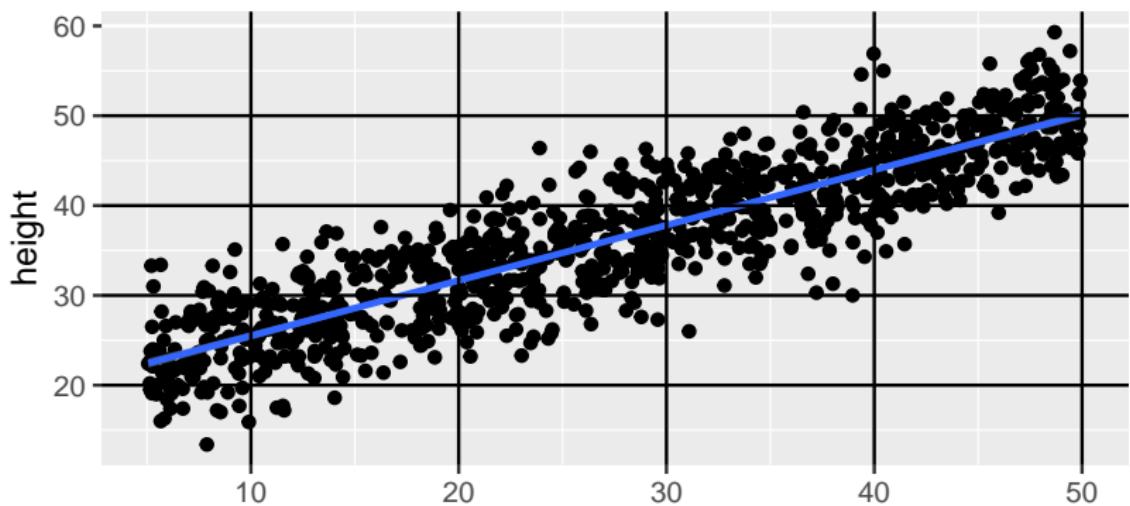
Adding more layers

```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point() +  
  geom_smooth(method = "lm") +  
  geom_vline(xintercept = c(10, 20, 30, 40, 50))
```



Adding more layers

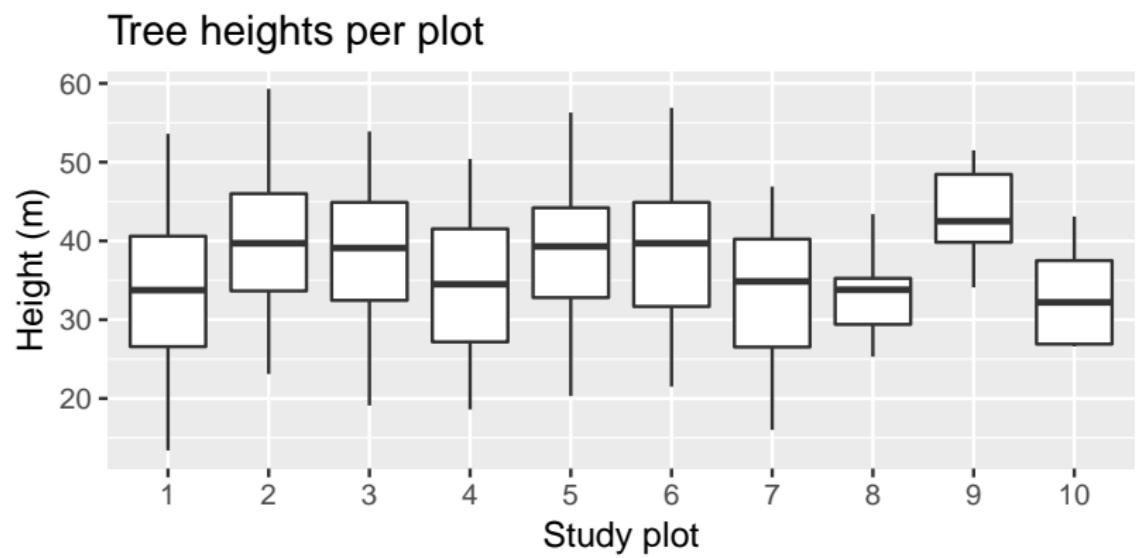
```
ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point() +  
  geom_smooth(method = "lm") +  
  geom_vline(xintercept = c(10, 20, 30, 40, 50)) +  
  geom_hline(yintercept = c(20, 30, 40, 50))
```



Summary

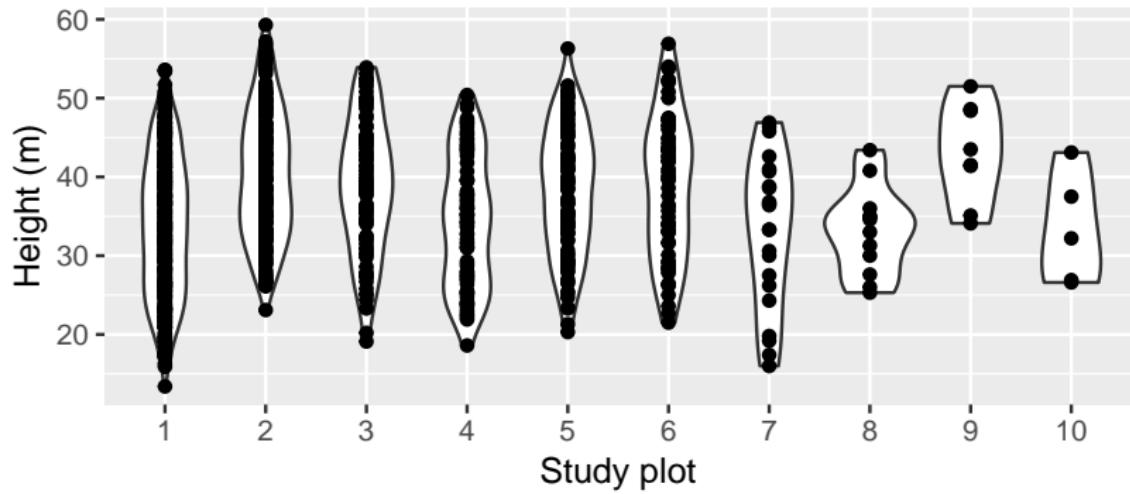
```
ggplot(trees) +          # Name of (tidy) data frame  
  aes(x = dbh, y = height) + # Aesthetics (variables to map in axes)  
  geom_point()             # Geoms: geometric objects
```

Exercise: Make a plot like this one



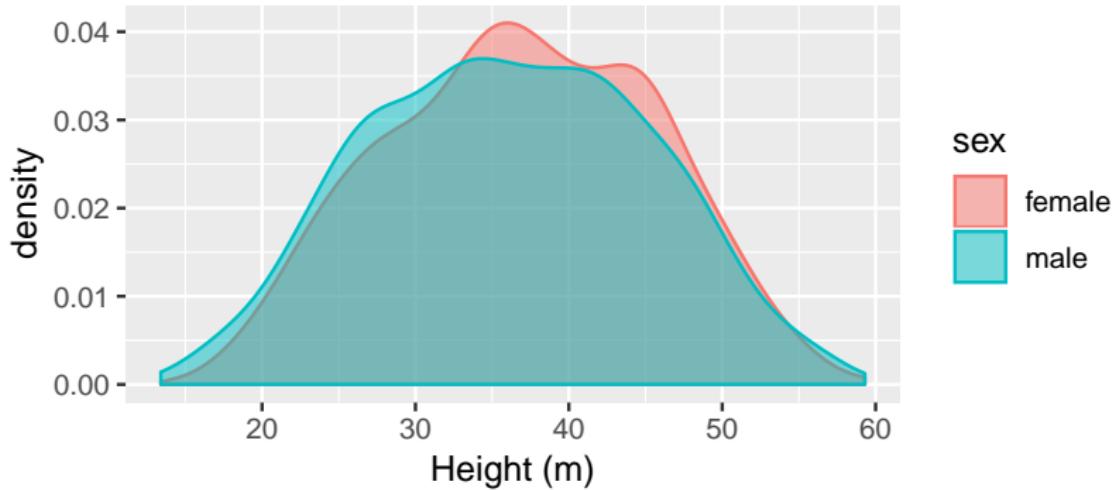
Exercise: Make a plot like this one

Tree heights per plot



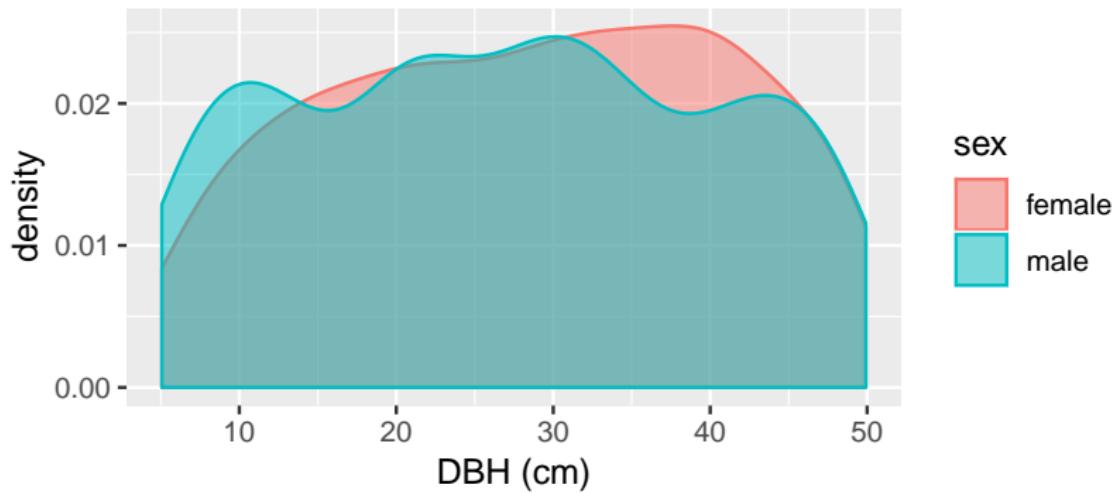
Exercise: Make a plot like this one

Distribution of heights

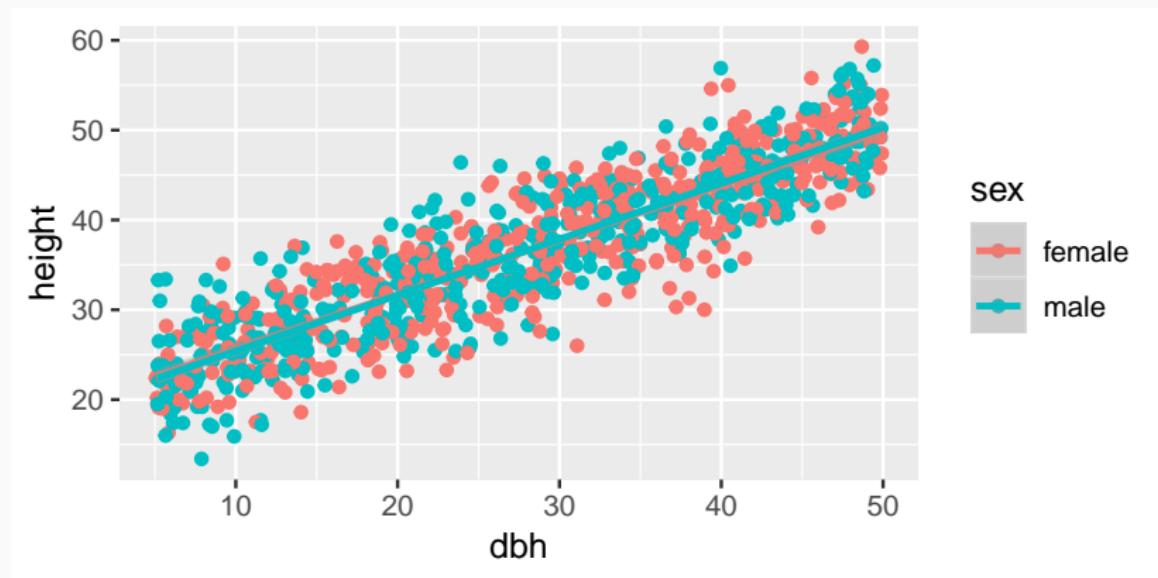


Exercise: Make a plot like this one

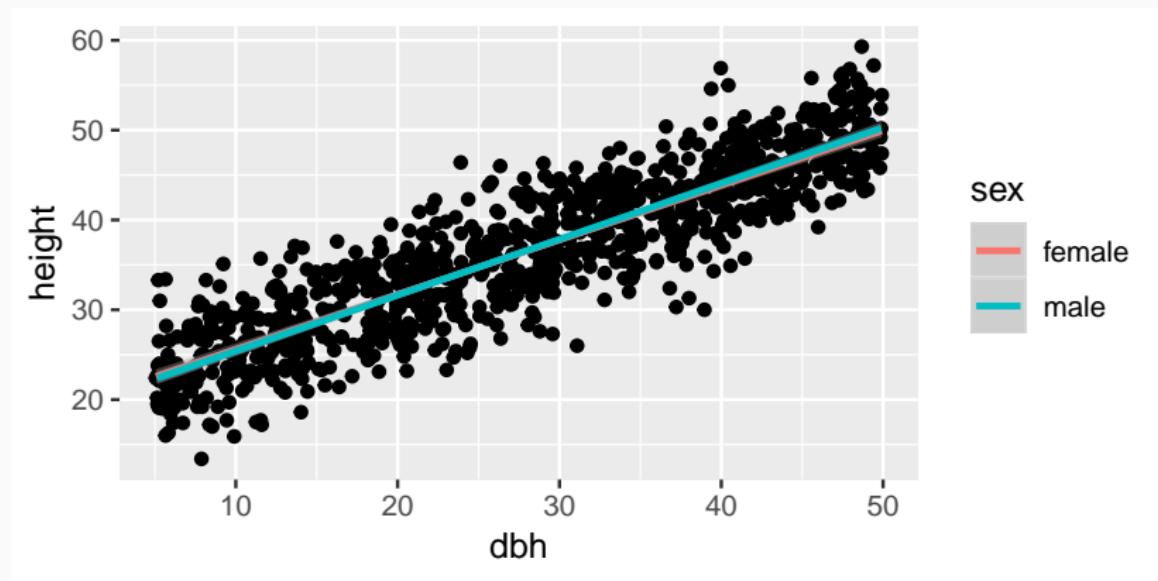
Distribution of diameters



Exercise: Make a plot like this one



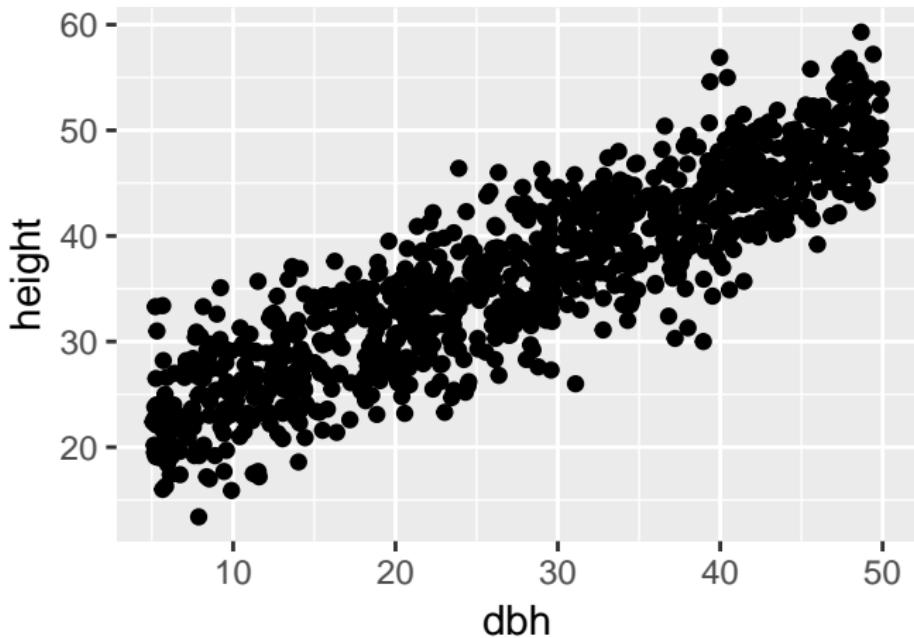
Exercise: Make a plot like this one



**ggplot2 figures can be assigned to R
objects**

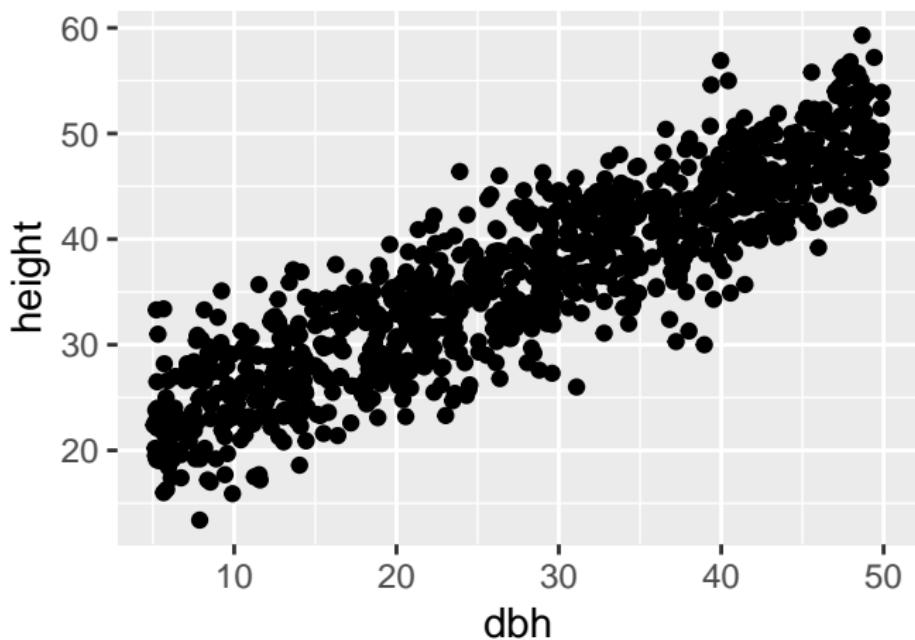
Assigning ggplot objects

```
myplot <- ggplot(trees) +  
  aes(x = dbh, y = height)  
myplot + geom_point()
```



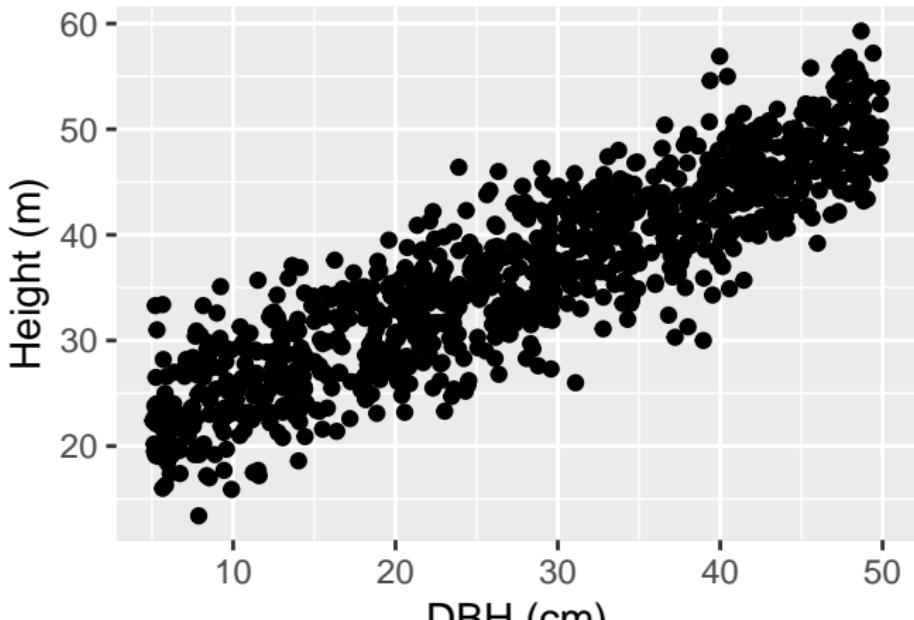
Assigning ggplot objects

```
myplot <- ggplot(trees) +  
  aes(x = dbh, y = height)  
myplot <- myplot + geom_point()  
myplot
```



Assigning ggplot objects

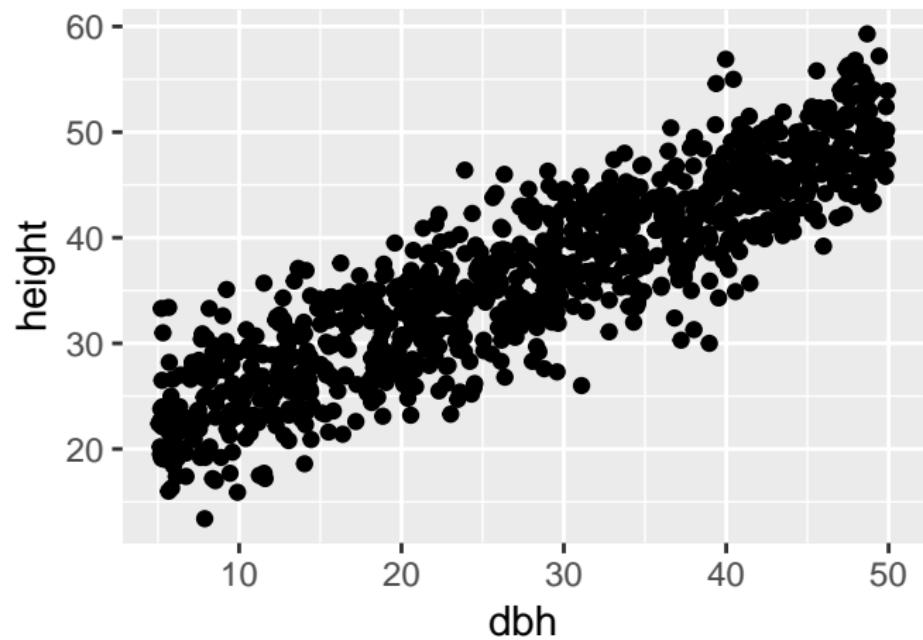
```
baseplot <- ggplot(trees) +  
  aes(x = dbh, y = height)  
scatterplot <- baseplot + geom_point()  
labelled <- scatterplot + labs(x = "DBH (cm)", y = "Height (m)")  
labelled
```



Themes: changing plot appearance

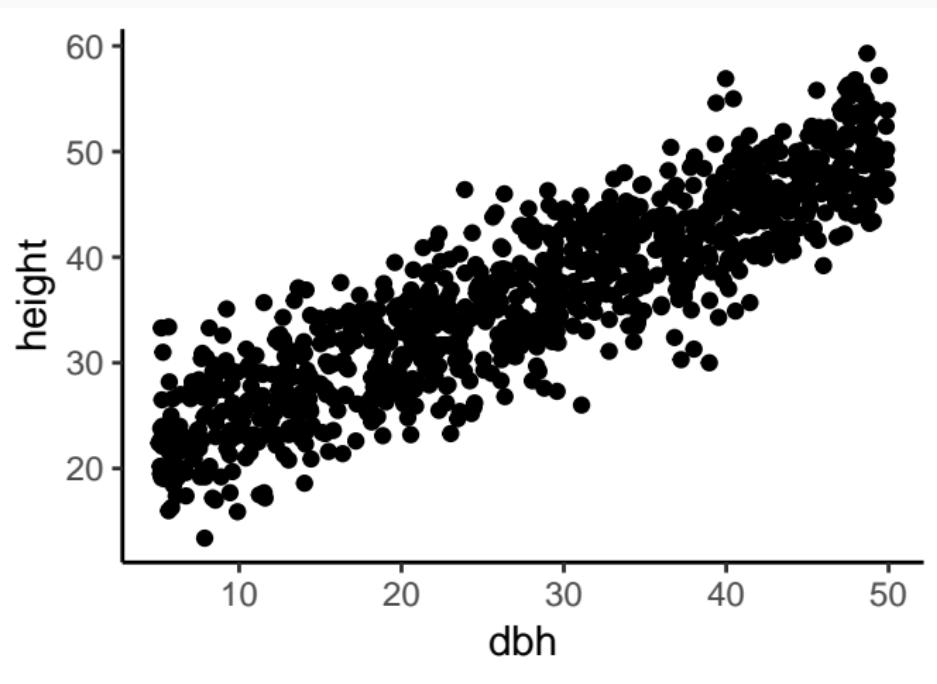
myplot

```
myplot <- ggplot(trees) +  
  aes(x = dbh, y = height) +  
  geom_point()
```



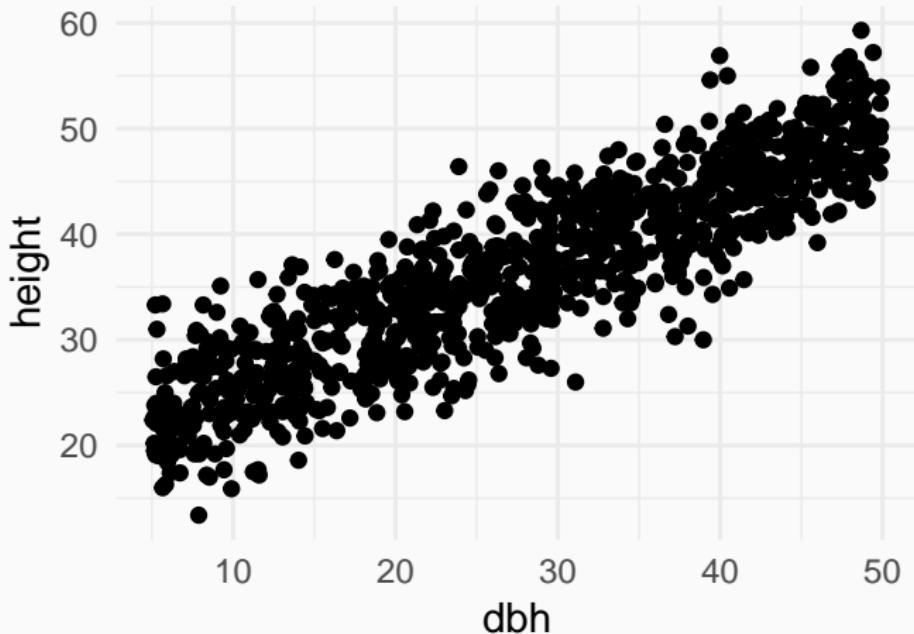
theme_classic

```
myplot + theme_classic()
```



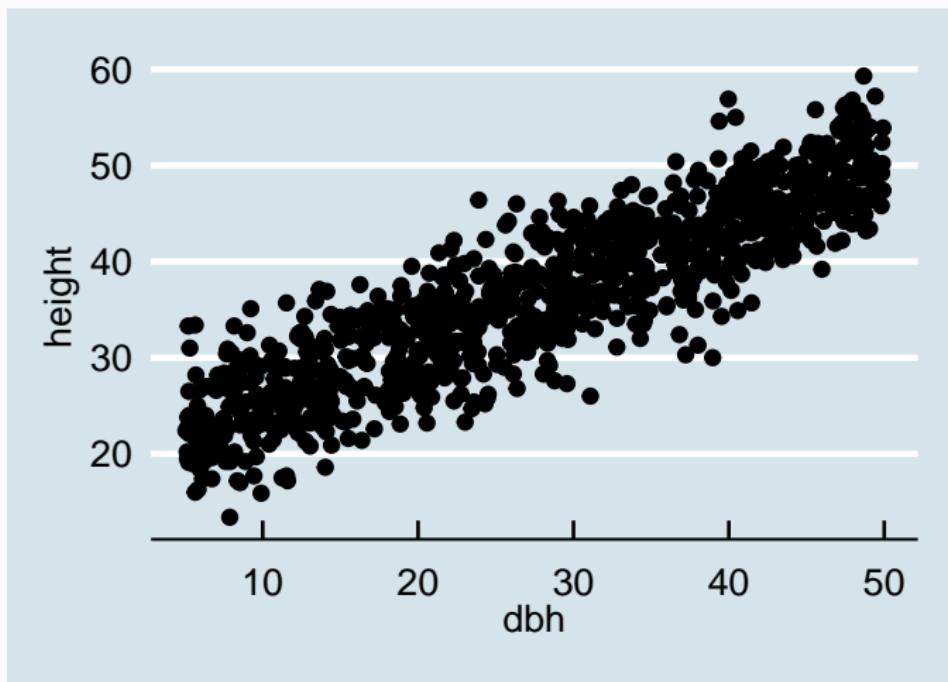
theme_minimal

```
myplot + theme_minimal()
```



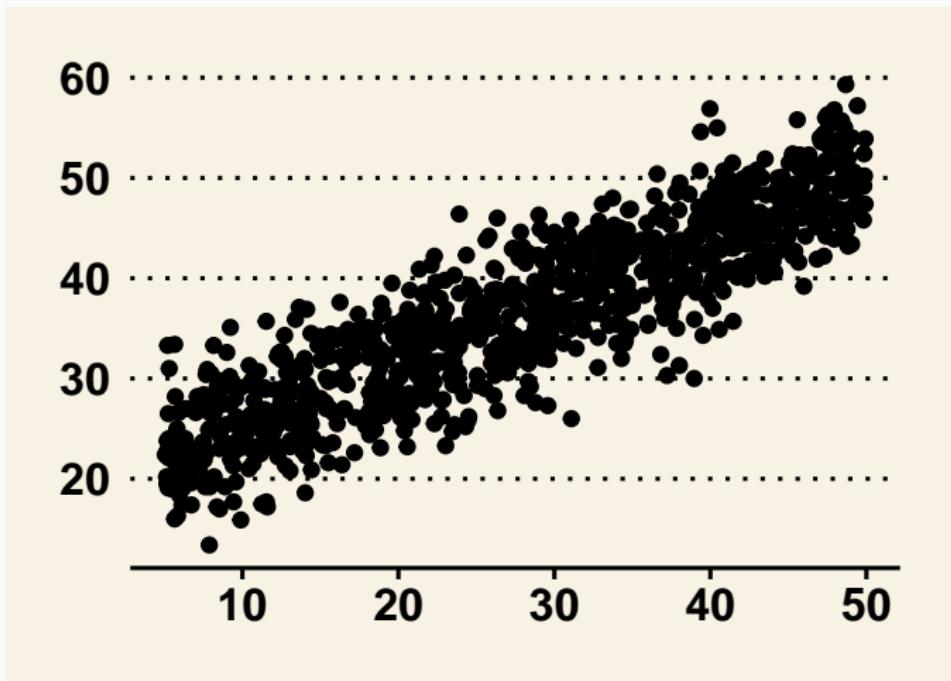
Lots of themes out there

```
library(ggthemes)  
myplot + theme_economist()
```



Lots of themes out there

```
myplot + theme_wsj()
```



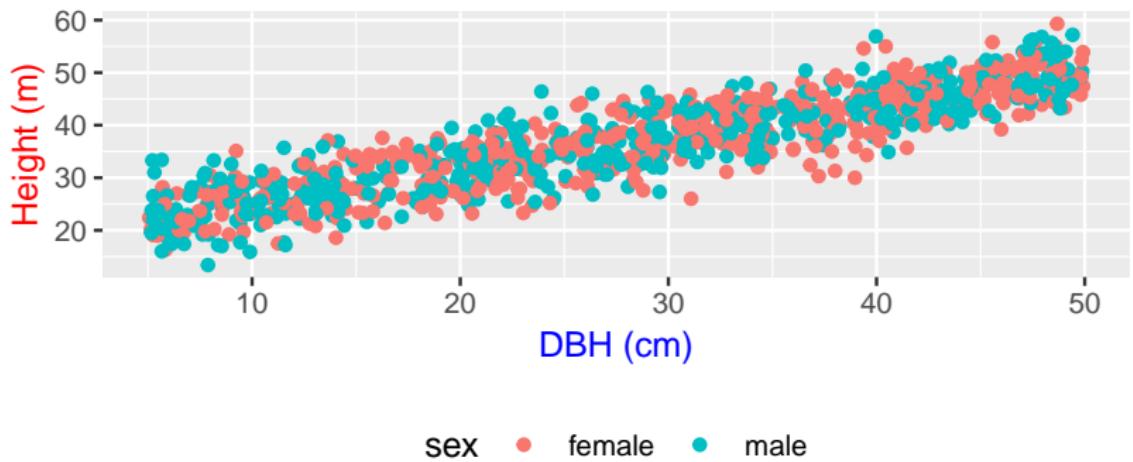
Editing themes

?theme

- element_blank
- element_text
- element_line
- element_rect (borders & backgrounds)

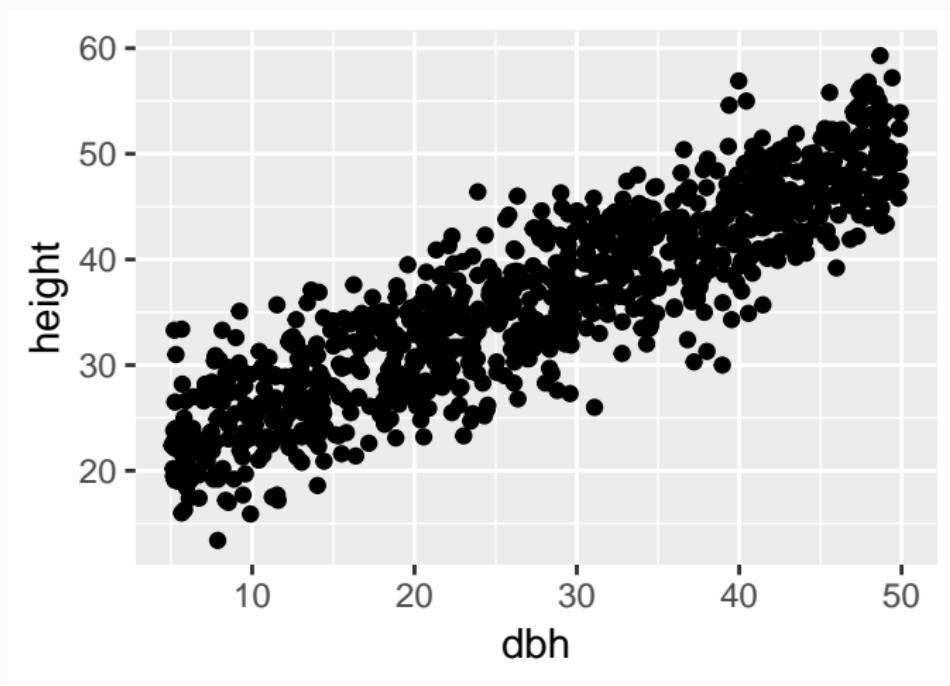
Exercise: make a plot like this one

Changing plot appearance



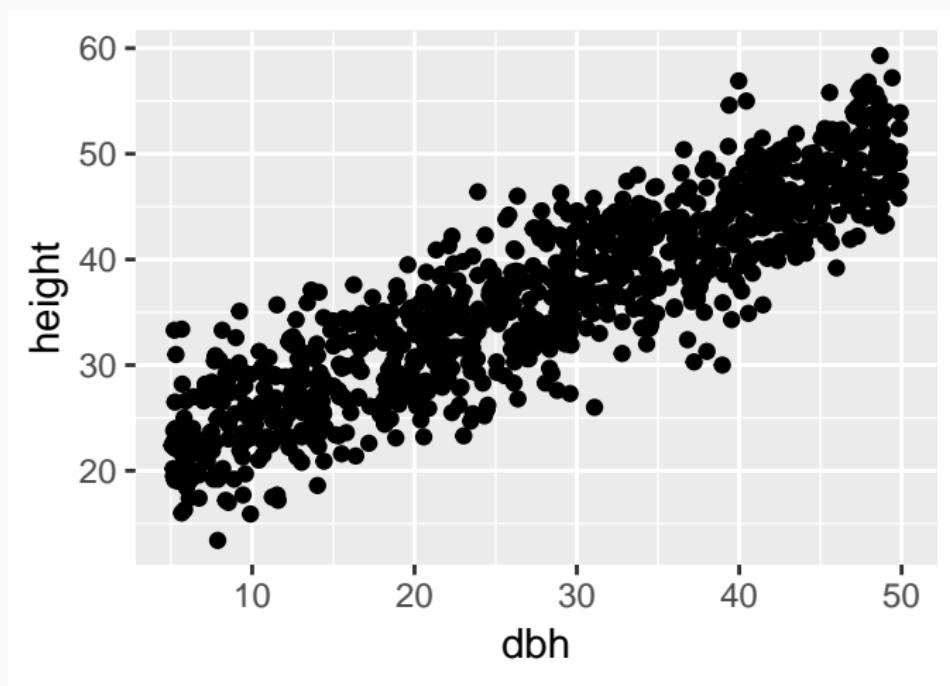
Easily changing appearance with ggthemeassist (Rstudio addin)

<https://github.com/calligross/ggthemeassist>



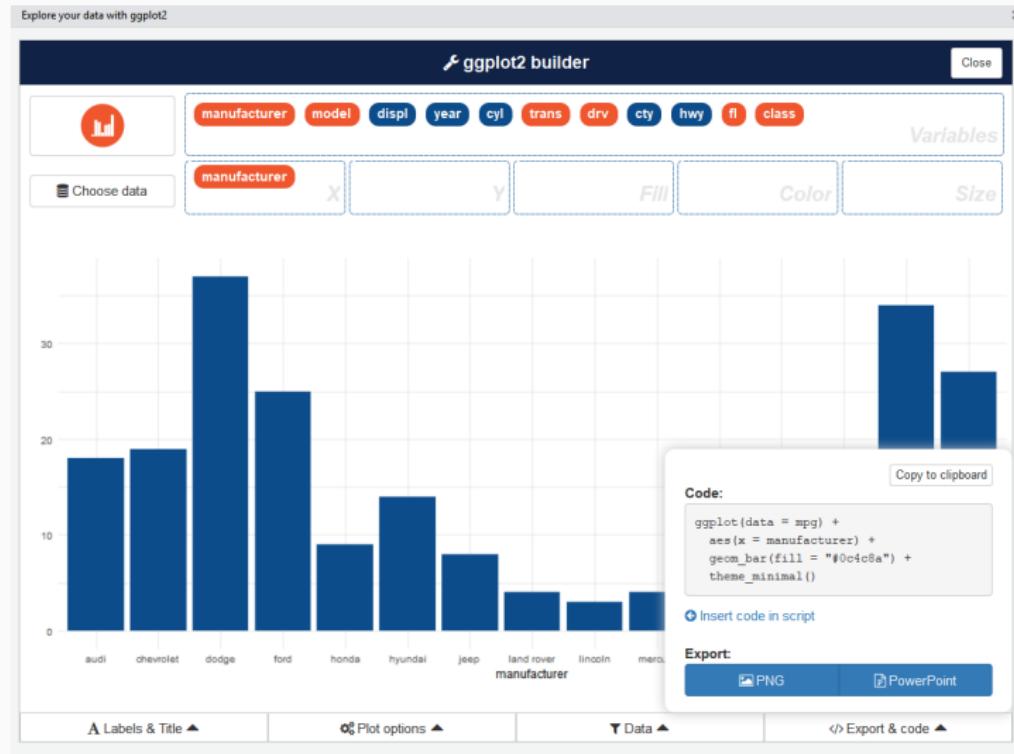
Easily changing appearance with ggedit

<https://github.com/metrumresearchgroup/ggedit>



esquisse: ggplot2 builder addin

<https://github.com/dreamRs/esquisse>





Trevor A. Branch
@TrevorABranch

Follow

My rule of thumb: every analysis you do on a dataset will have to be redone 10–15 times before publication. Plan accordingly. #Rstats

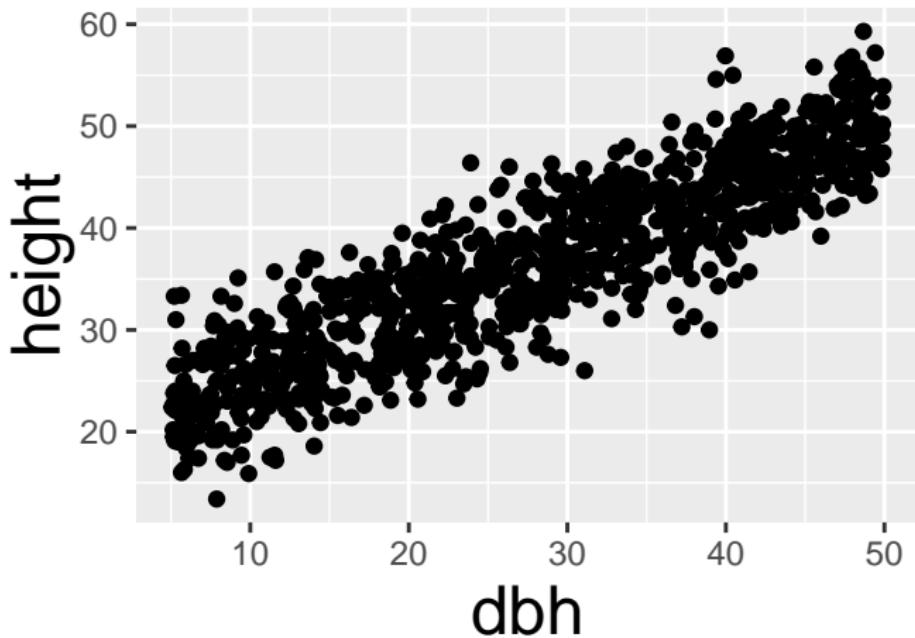
<http://mbjoseph.github.io/2015/02/26/plotting.html>

serialmentor.com/dataviz/choosing-the-right-visualization-software.html

Think twice before editing plots out of R

Referee #3: "Please increase font size in all figures"

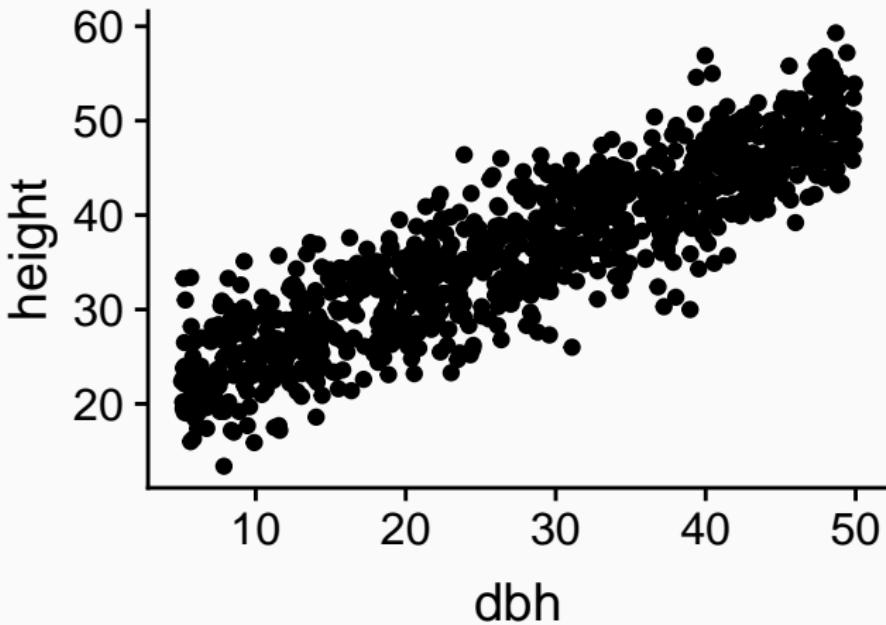
```
myplot +  
  theme(axis.title = element_text(size = 18))
```



Publication-quality plots

```
library(cowplot)
```

```
myplot
```



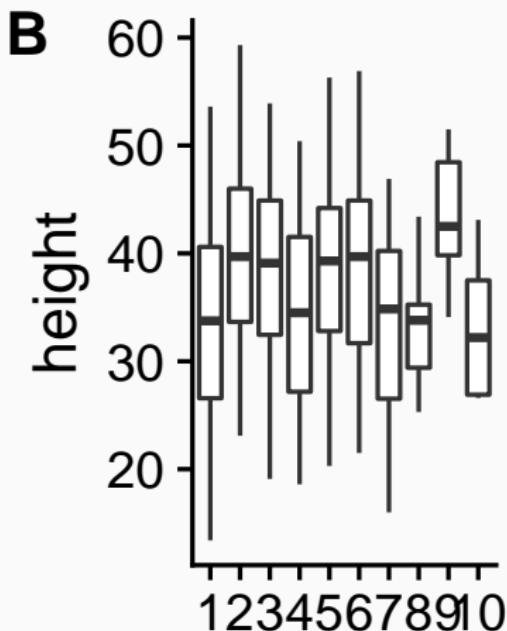
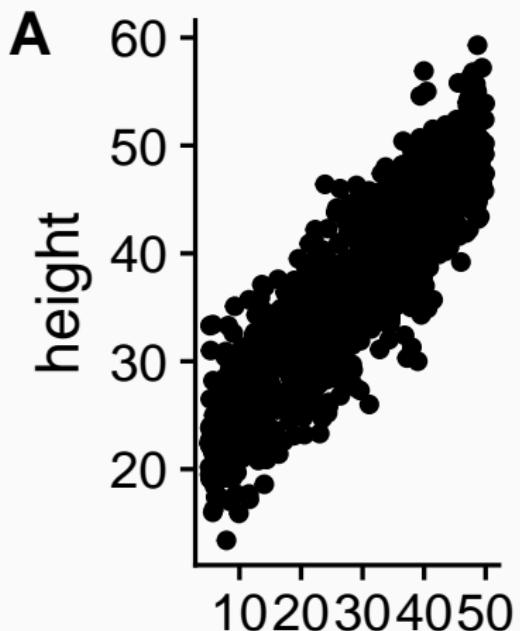
Some publication themes:

<https://gist.github.com/Pakillo/c2c7ea11c528cc2ee20f#themes>

Composite figures

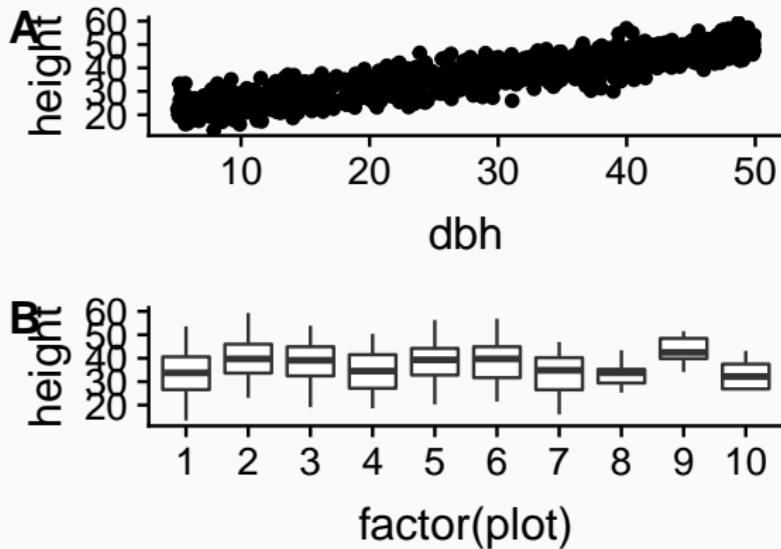
Composite figures: cowplot

```
library(cowplot)  
plot1 <- ggplot(trees) + aes(dbh, height) + geom_point()  
plot2 <- ggplot(trees) + aes(factor(plot), height) + geom_boxplot()  
plot_grid(plot1, plot2, labels = "AUTO")
```

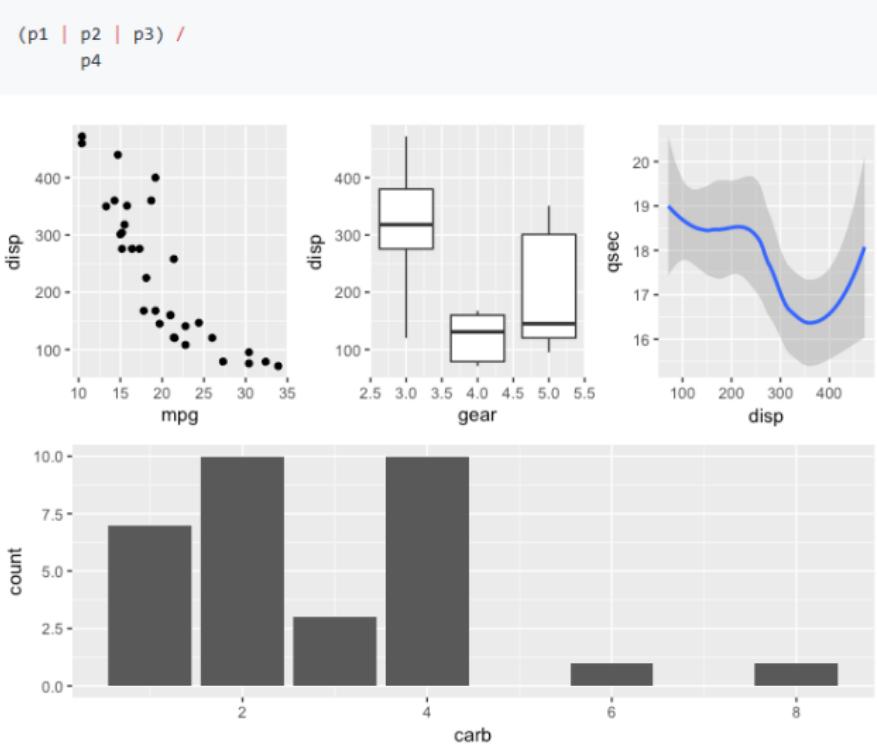


Composite figures

```
plot_grid(plot1, plot2, labels = "AUTO", ncol = 1)
```

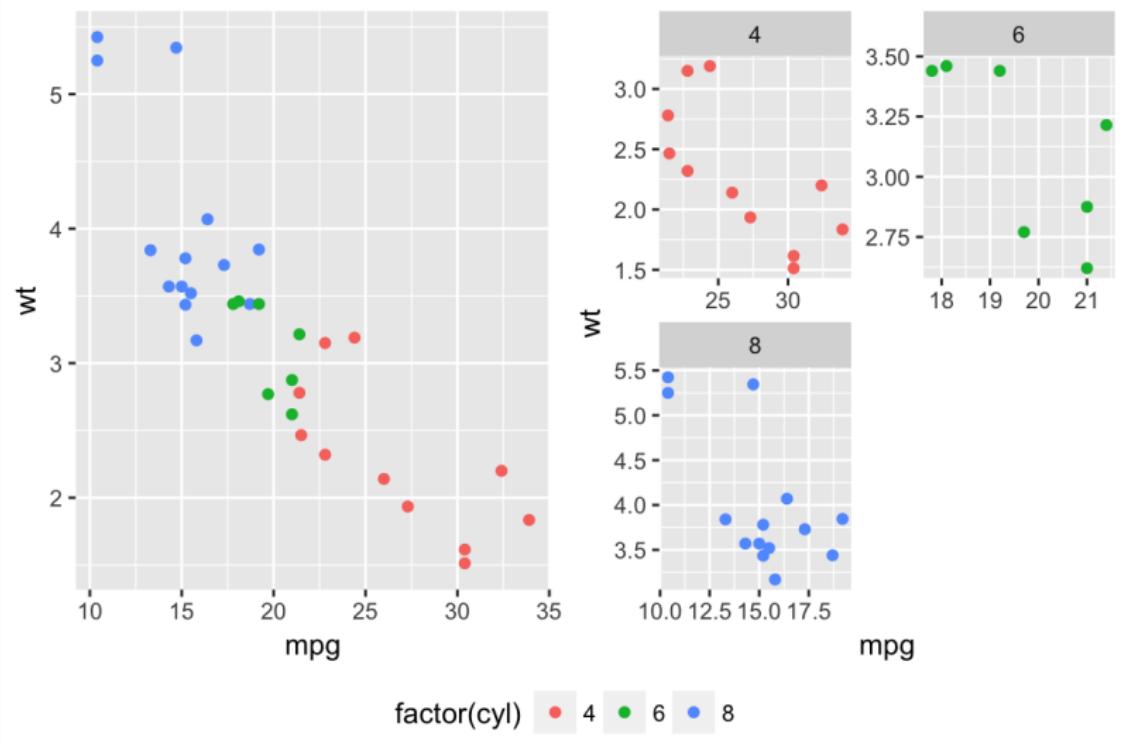


Composite figures: patchwork



<https://github.com/thomasp85/patchwork>

Composite figures: egg



<https://cran.r-project.org/web/packages/egg/index.html>

Saving plot

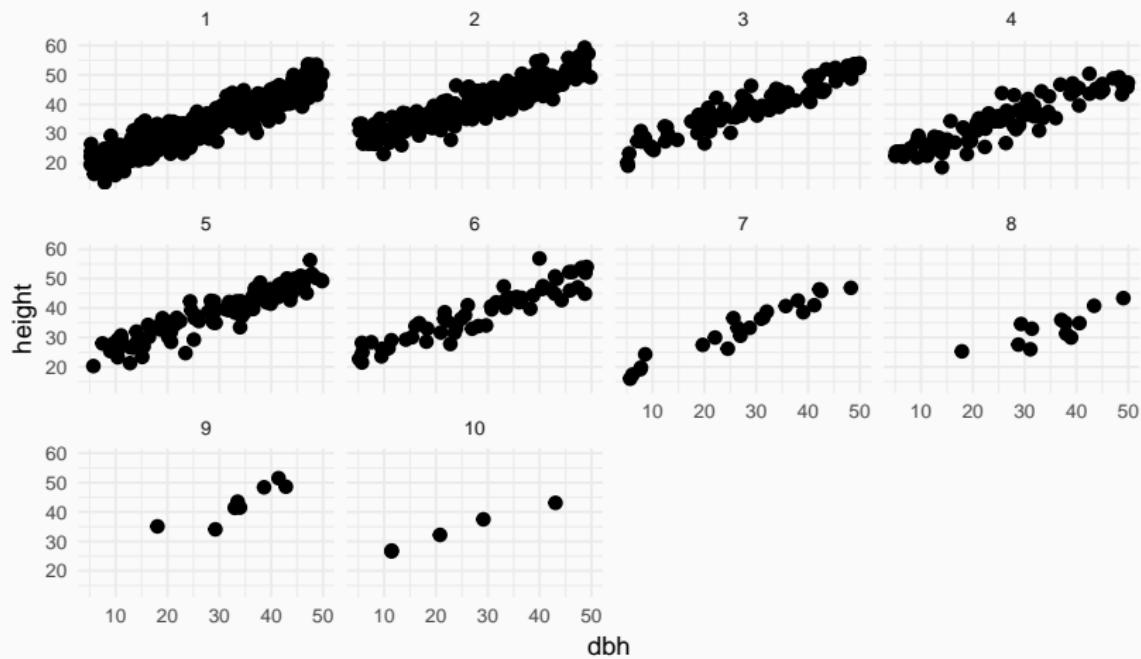
```
ggsave("myplot.pdf")
```

```
save_plot("myplot.pdf")
```

Facetting (small multiples)

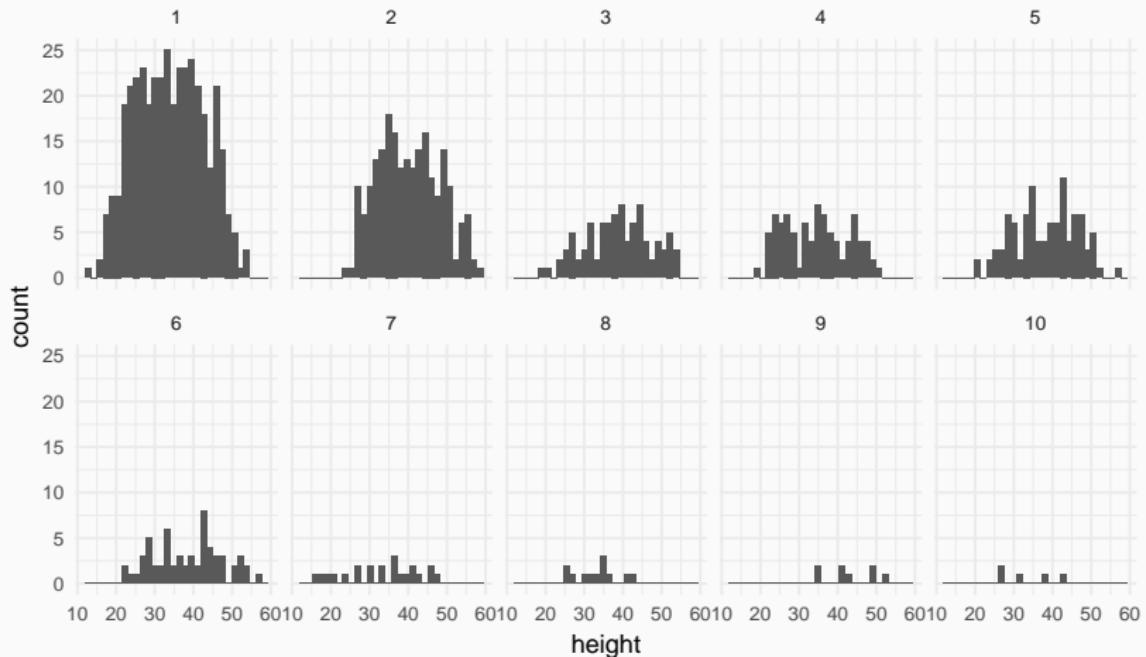
Facetting

```
ggplot(trees) + aes(dbh, height) +  
  geom_point() + theme_minimal(base_size = 8) +  
  facet_wrap(~plot)
```



Facetting

```
ggplot(trees) +  
  geom_histogram(aes(height)) + theme_minimal(base_size = 8) +  
  facet_wrap(~plot, nrow = 2)
```

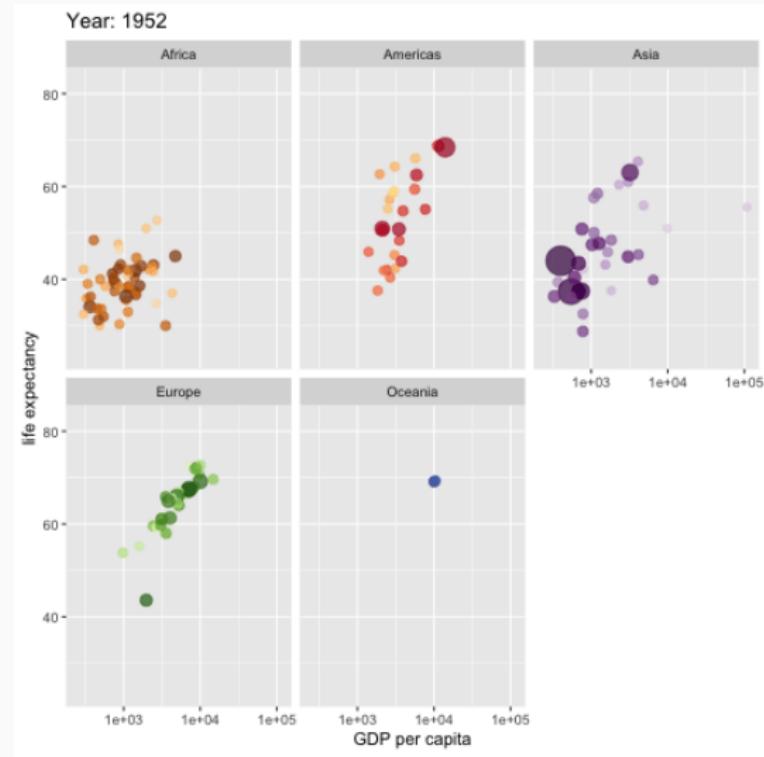


Interactivity: plotly

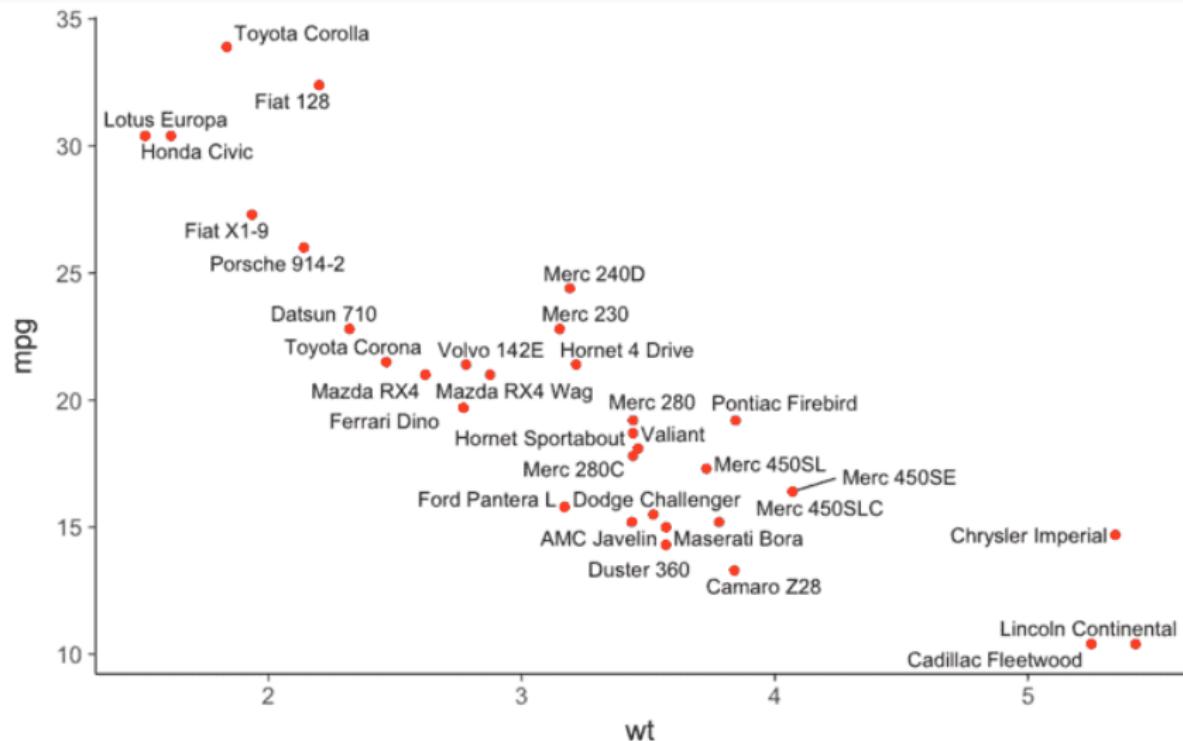
```
library(plotly)
myplot <- ggplot(trees) +
  aes(x = dbh, y = height) +
  geom_point()
ggplotly(myplot)
```

Animated graphs

<https://github.com/thomasp85/gganimate>



Automatic label placement



Many extensions!

<https://www.ggplot2-exts.org/>

40 registered extensions available to explore

Sort: Github stars Filter: search name, author, desc... Author Filter Tag Filter CRAN Only

Showing 40 of 40

gg3D (star)
3D perspective plots for ggplot2.
• **author:** Daniel Adler
• **tags:** 3D, Visualization
• **js libraries:**

ggQC (star)
Use ggQC to plot single, faceted and multi-layered quality control charts.
• **author:** Kenneth Gray
• **tags:** QC, XbarR, XbarS, SixSigma, Visualization
• **js libraries:**

ggedit (star)
ggedit is aimed to interactively edit ggplot layers, scales and themes aesthetics
• **author:** jordic
• **tags:** visualization, interactive, shiny, general
• **js libraries:**

ggforce (star)
ggforce is aimed at providing missing functionality to ggplot2 through the extension system introduced with ggplot2 v2.0.0.
• **author:** thomasp85
• **tags:** visualization, general
• **js libraries:**

ggalt (star)
A compendium of 'geom's, 'coords' and 'stats' for 'ggplot2'.
• **author:** hrbrmstr
• **tags:** visualization, general
• **js libraries:**

ggraph (star)
Homologous to make...
• **author:** denigrider
• **tags:** visualization, general
• **js libraries:**

ggstance (star)
ggstance implements horizontal versions of common ggplot2 geoms.
• **author:** forbesl
• **tags:** visualization, general
• **js libraries:**

ggrepel (star)
label overlapping text labels away from each other.
• **author:** slowflow
• **tags:** visualization, general
• **js libraries:**

ggraph (star)
ggraph is tailored at plotting graph-like data structures (graphs, networks, trees, hierarchies,...).
• **author:** thomasp85
• **tags:** visualization, general
• **js libraries:**

ggpmisc (star)
Miscellaneous Extensions to 'ggplot2'.
• **author:**
• **tags:** visualization, general
• **js libraries:**

geomnet (star)
geomnet implements network visualizations in ggplot2 via geom.net.
• **author:** schmitt
• **tags:** visualization, general
• **js libraries:**

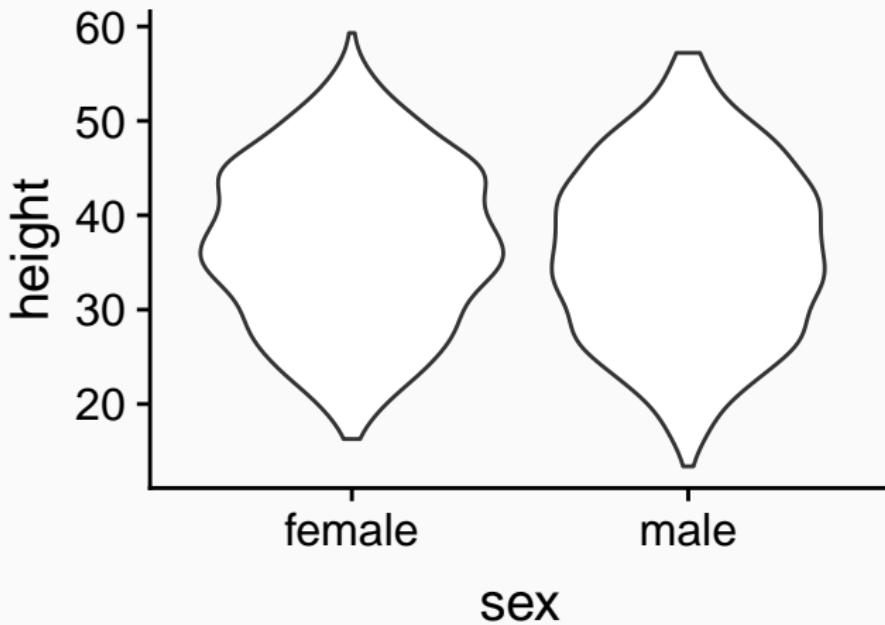
ggExtra (star)
ggExtra lets you add histograms to ggplot2.
• **author:** damian
• **tags:** visualization, general
• **js libraries:**

Summary

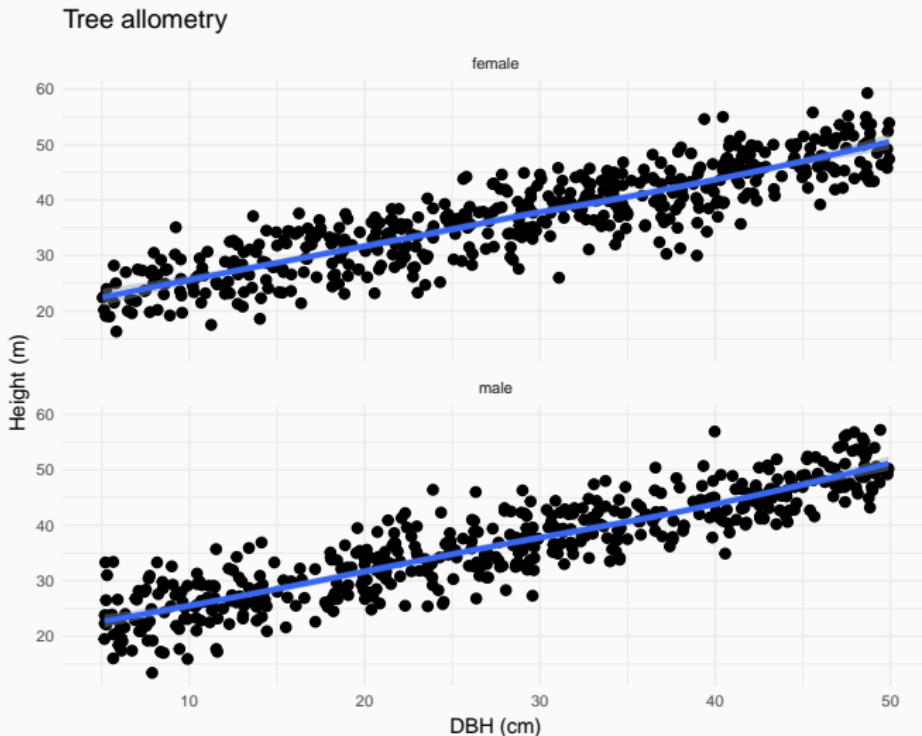
Grammar of graphics

- **Data** (tidy data frame)
- **Layers** (*geoms*: points, lines, polygons...)
- **Aesthetics** mappings (x, y, size, colour...)
- **Scales** (colour, size, shape...)
- **Facets** (small multiples)
- **Themes** (appearance)
- **Coordinate system** (Cartesian, polar, map projections...)

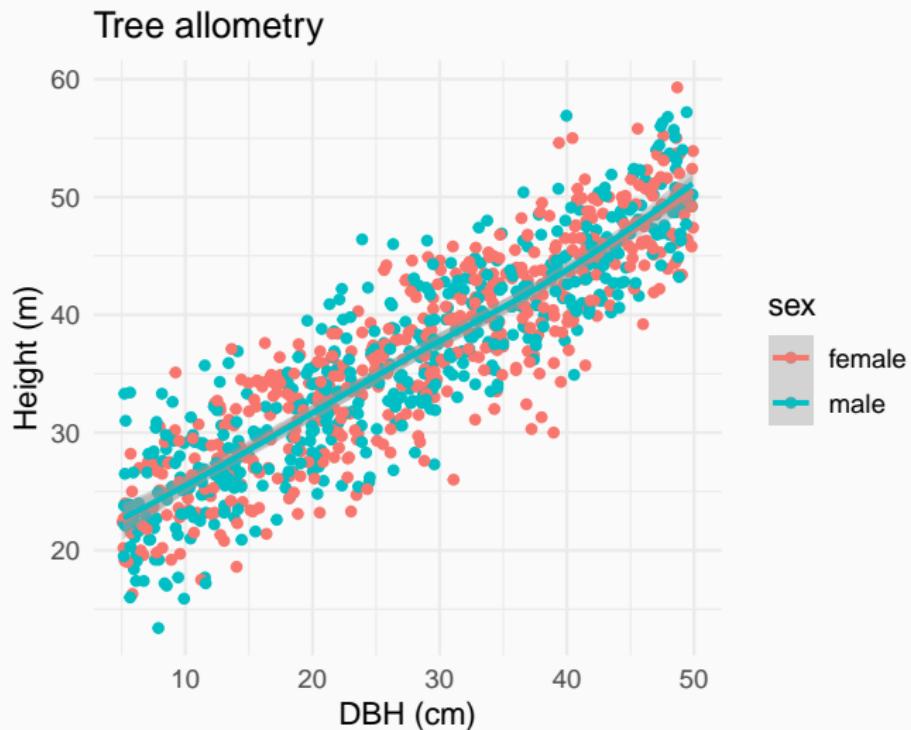
Exercise: make a plot like this one



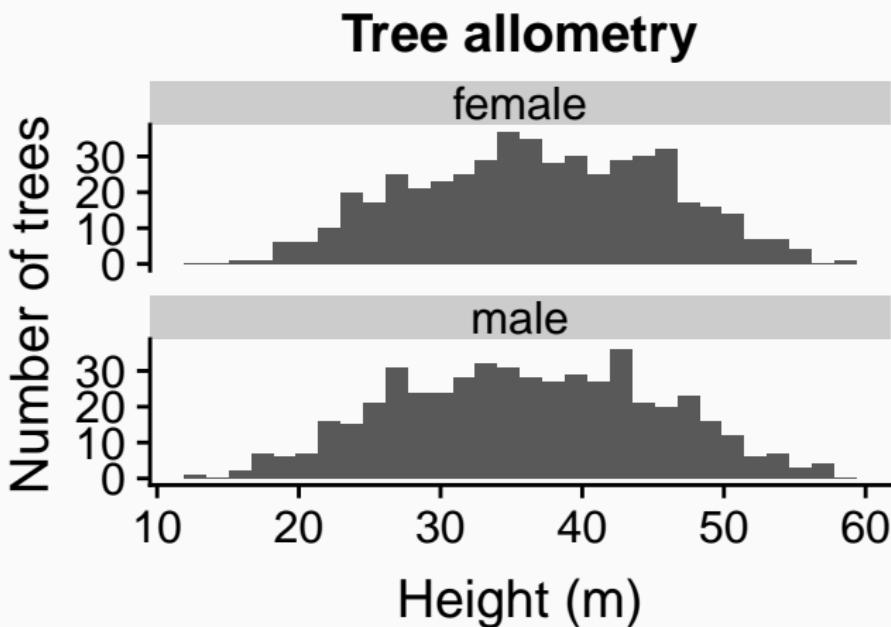
Exercise: make a plot like this one



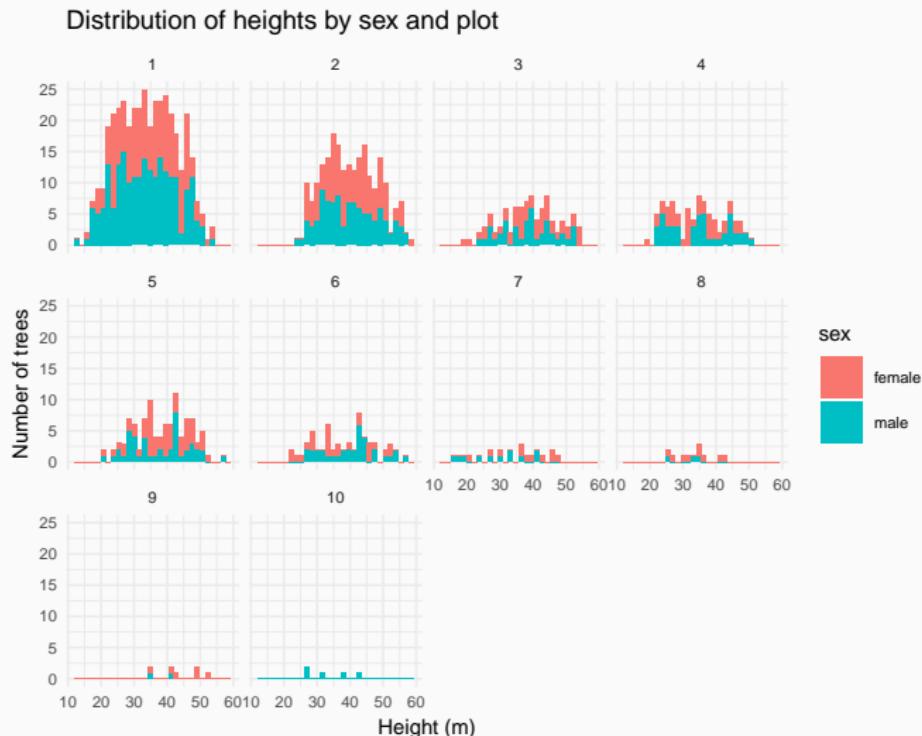
Exercise: make a plot like this one



Exercise: make a plot like this one

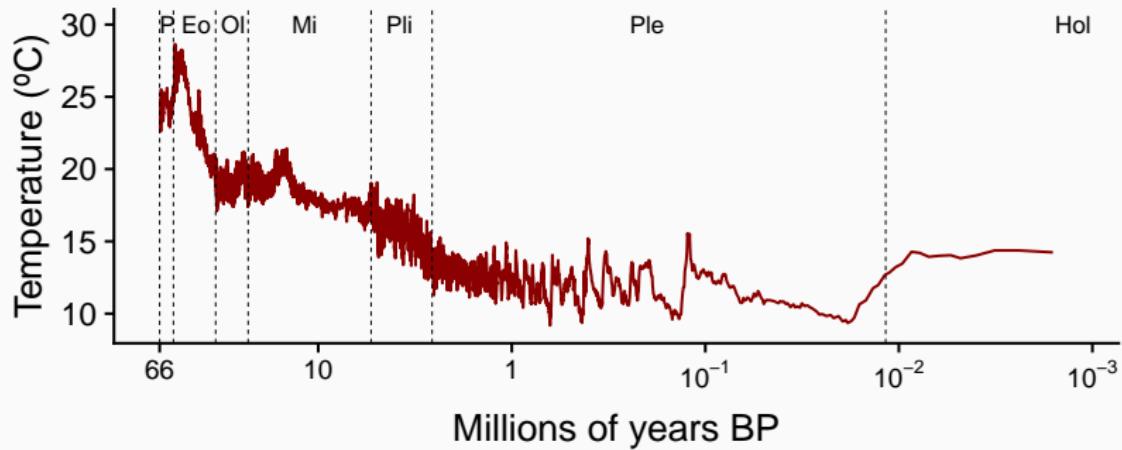


Exercise: make a plot like this one

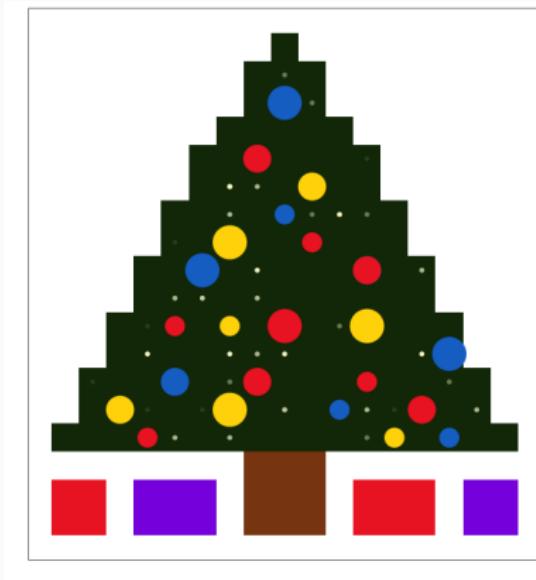


Exercise: make a plot like this one

Data from <http://www.columbia.edu/~mhs119/Sensitivity+SL+CO2/Table.txt>



Exercise: make a plot like this one



END



Slides and source code available at <https://github.com/Pakillo/ggplot-intro>