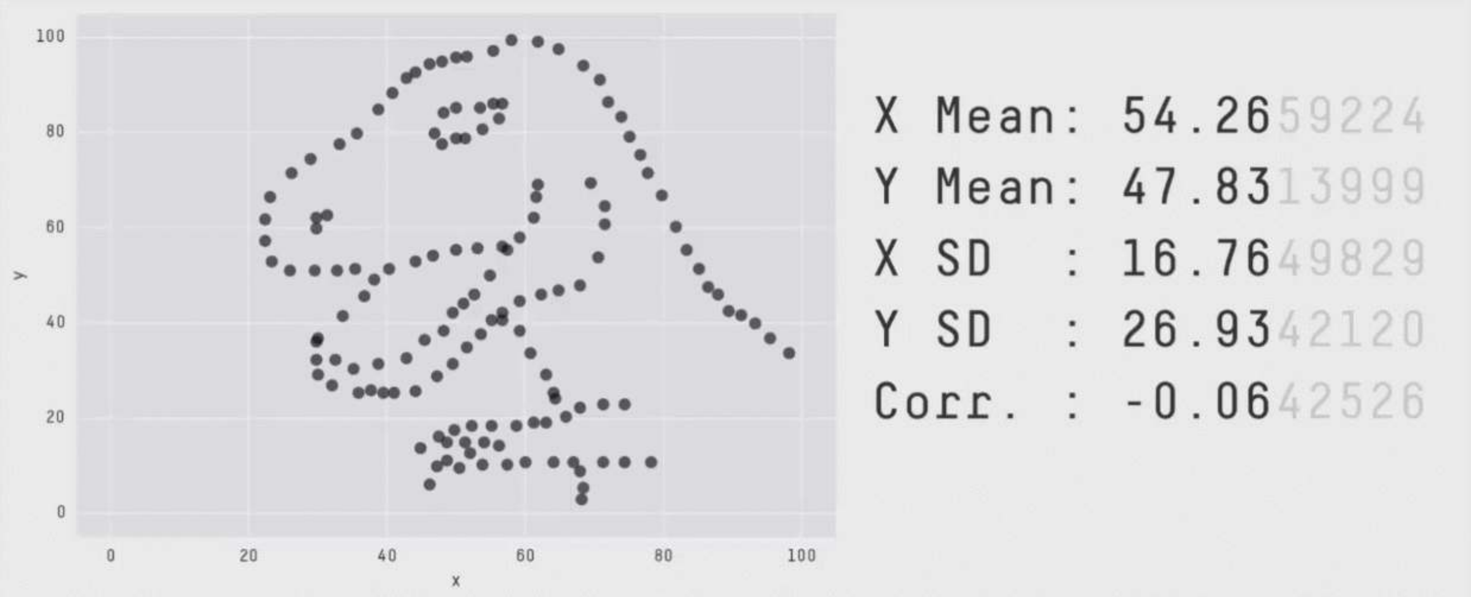
Data visualisation with ggplot2

Francisco Rodríguez-Sánchez

https://frodriguezsanchez.net

@frod_san

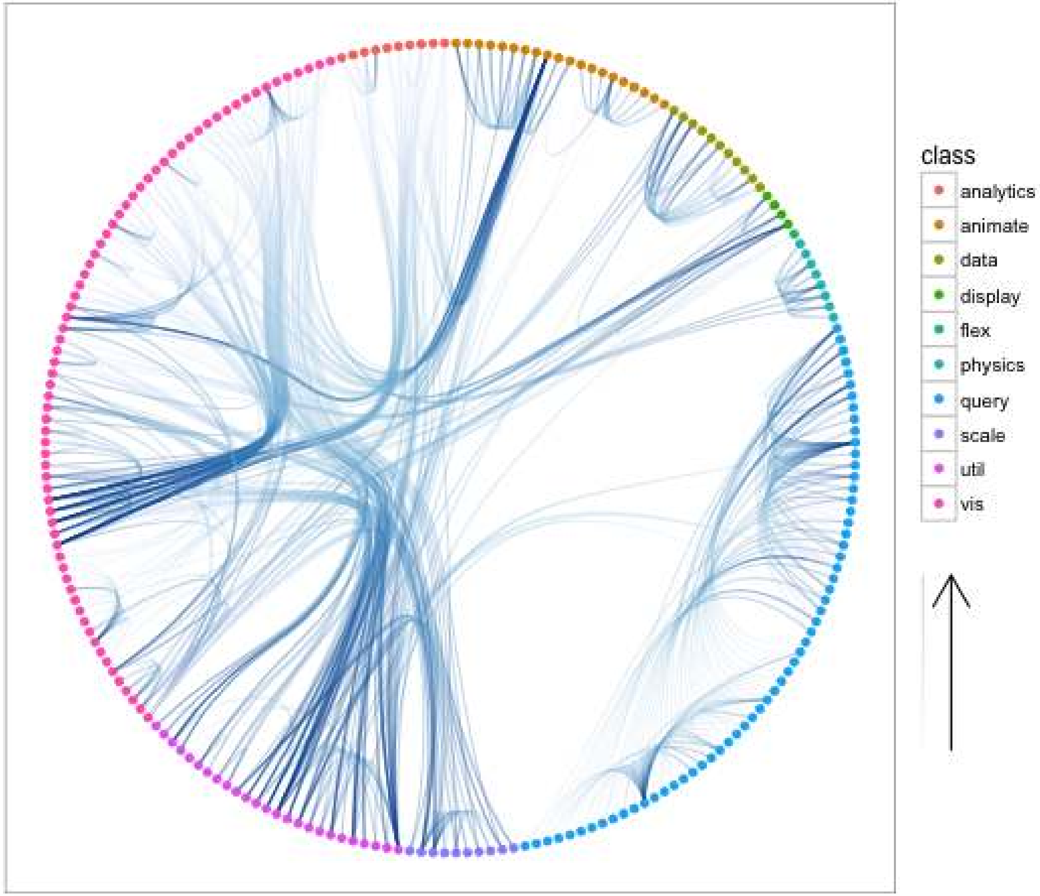
Always plot data!



```
X Mean: 54.2659224
Y Mean: 47.8313999
X SD  : 16.7649829
Y SD  : 26.9342120
Corr. : -0.0642526
```

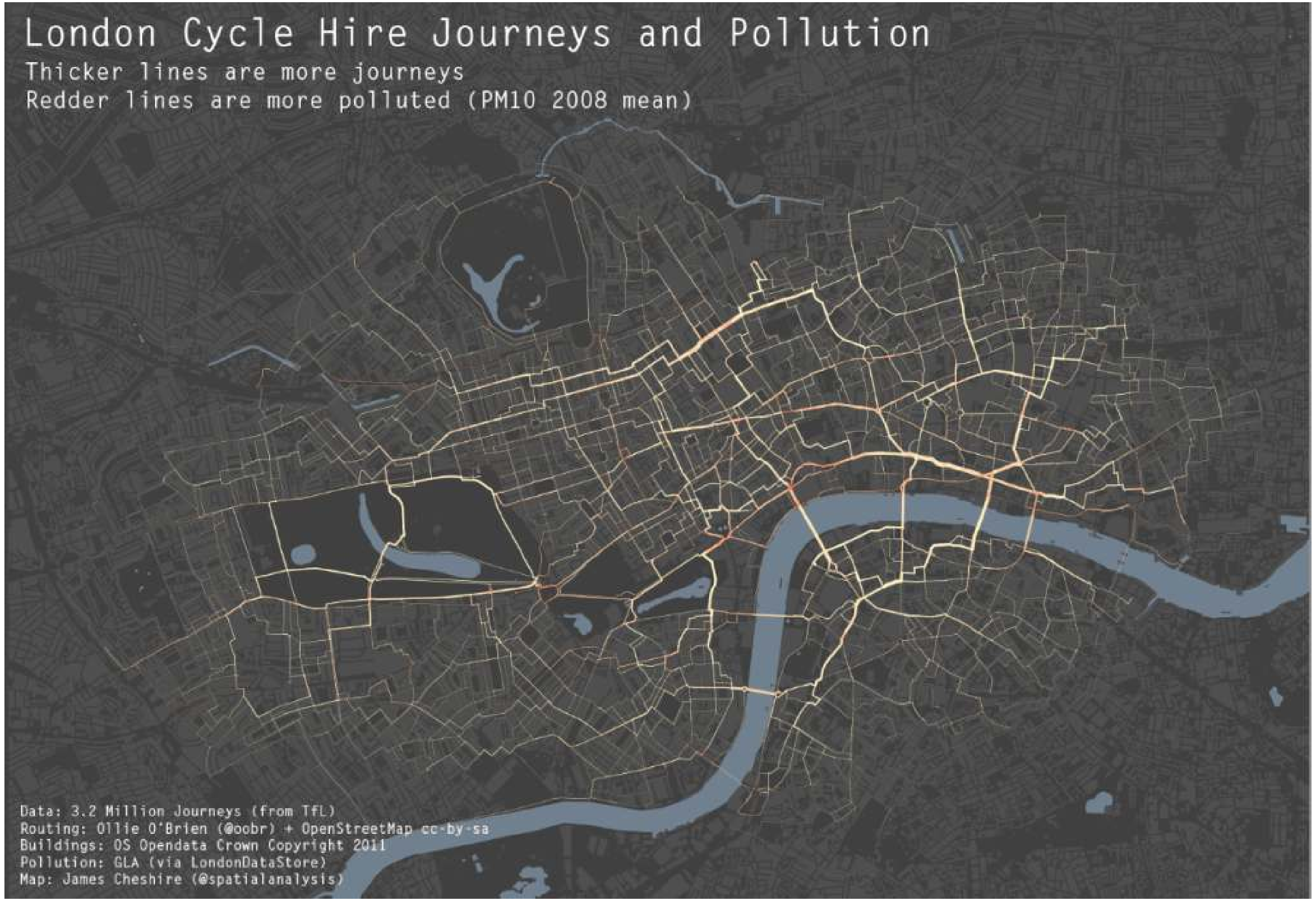
https://github.com/stephlocke/datasauRus

Made with ggplot



https://github.com/thomasp85/ggraph

Made with ggplot



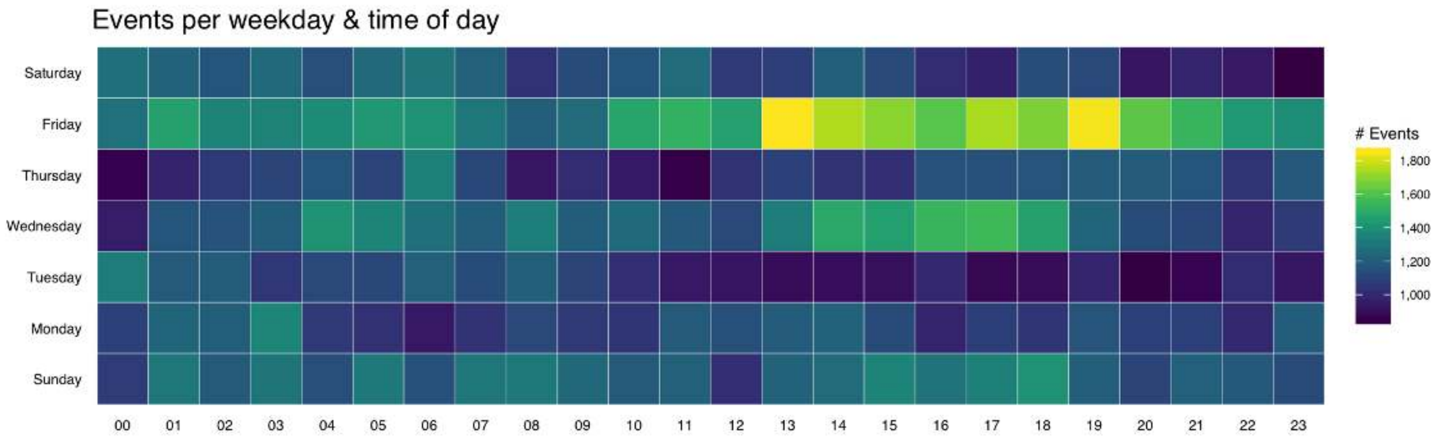
London Cycle Hire Journeys and Pollution

Thicker lines are more journeys
Redder lines are more polluted (PM10 2008 mean)

Data: 3.2 Million Journeys (from TfL)
Routing: Ollie O'Brien (@oobr) + OpenStreetMap cc-by-sa
Buildings: OS Opendata Crown Copyright 2011
Pollution: GLA (via LondonDataStore)
Map: James Cheshire (@spatialanalysis)

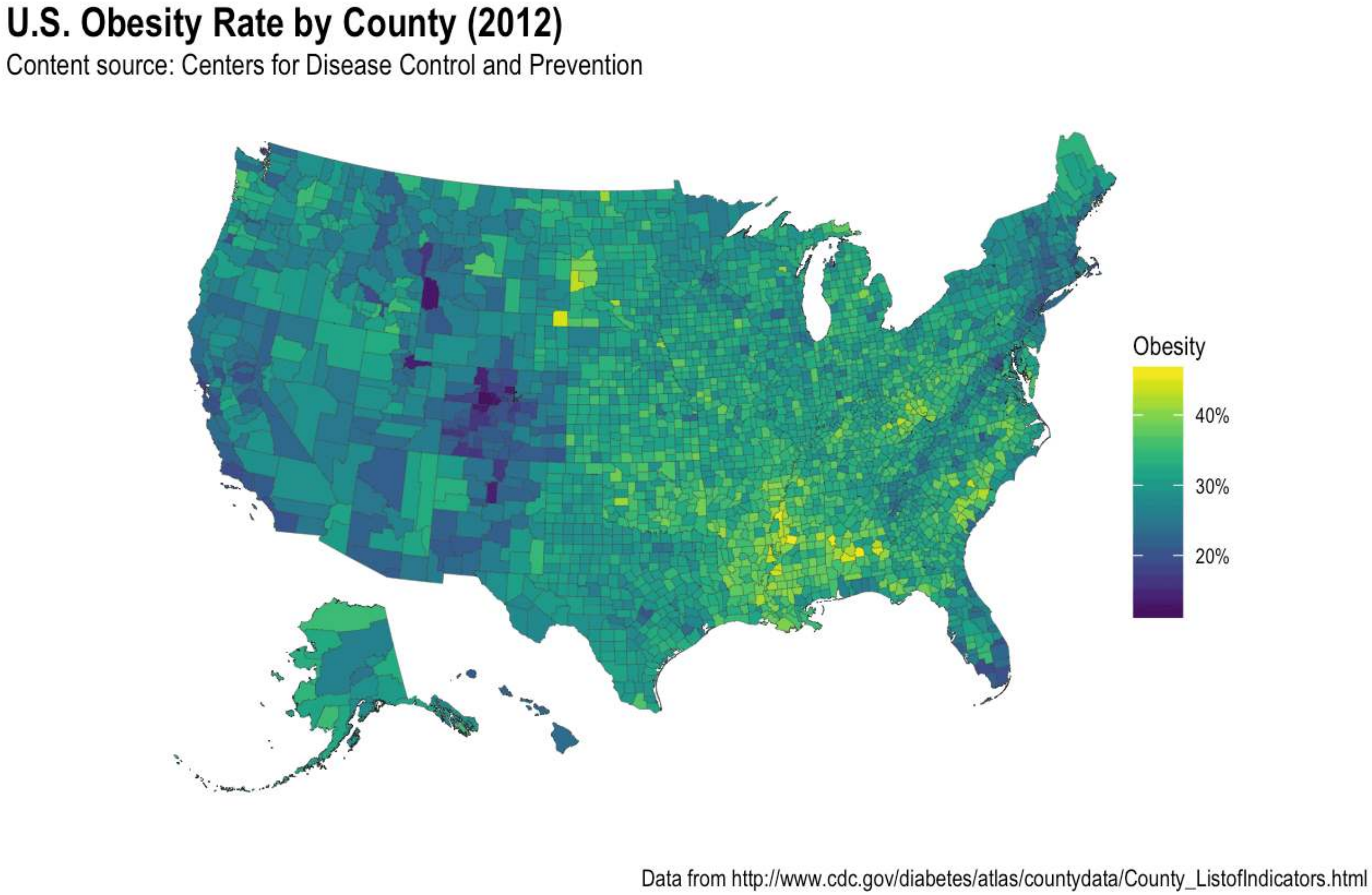
http://spatial.ly/2012/02/great-maps-ggplot2/

Made with ggplot



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

Made with ggplot



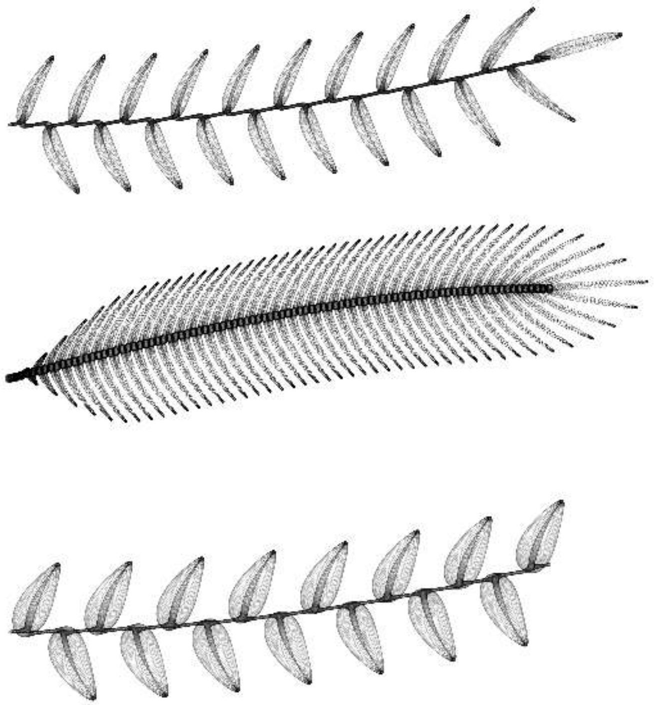
U.S. Obesity Rate by County (2012)

Content source: Centers for Disease Control and Prevention

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

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

Made with ggplot



https://github.com/marcusvolz/mathart

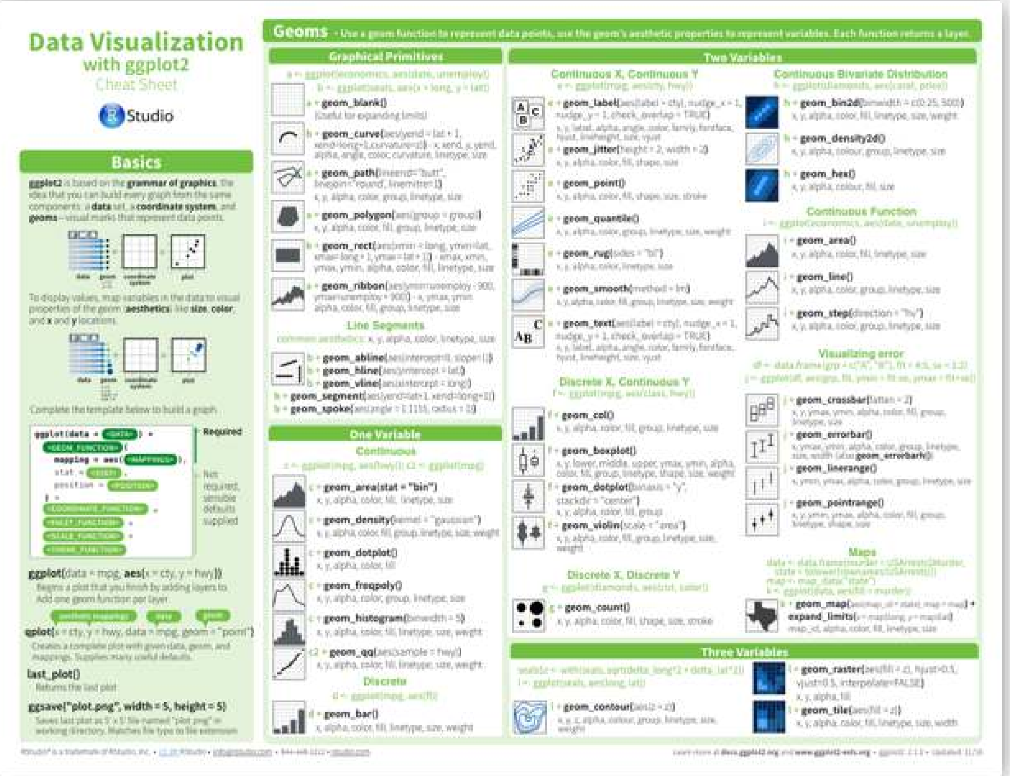
Why ggplot

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

Very good documentation and tutorials

- Official ggplot2 documentation
- ggplot2 book
- Data visualisation chapter in R for Data Science
- R graphics cookbook and Cookbook for R
- Data visualization: a practical introduction (K. Healy)
- Fundamentals of data visualization (C. Wilke)

Cheatsheet



Data Visualization with ggplot2 Cheat Sheet

Studio

Basics

Geoms

Graphical Primitives

Line Segments

One Variable

Continuous

Discrete

Two Variables

Continuous X, Continuous Y

Discrete X, Continuous Y

Discrete X, Discrete Y

Continuous Bivariate Distribution

Continuous Function

Visualizing error

Maps

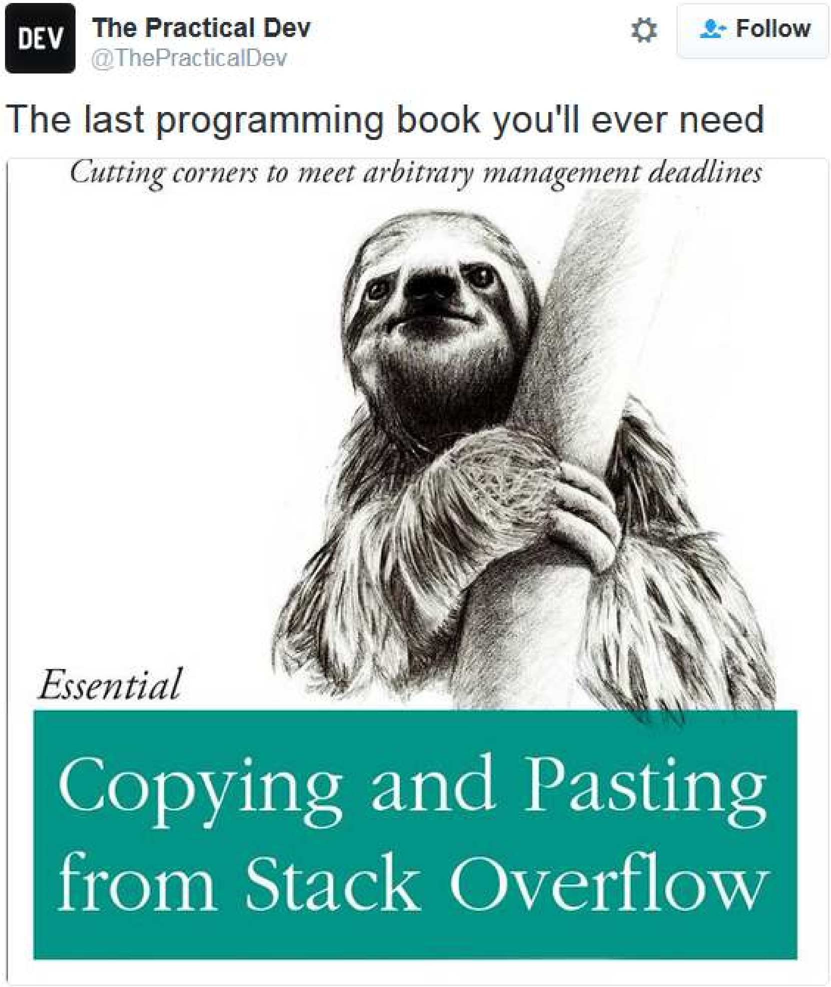
Three Variables

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

Repos of figures + code

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

Find answers in Stack Overflow, Rstudio Community, R4DS...



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: paper planes flying experiment

```r
library(paperplanes)
data(paperplanes)
head(paperplanes)
```

```
# A tibble: 6 × 8
     id hour    person   gender   age plane       paper distance
  <int> <fct>   <chr>    <fct>  <dbl> <chr>       <int>    <dbl>
1     1 [17,18) Roland   male      30 Standard80     80      7.8
2     2 [17,18) Astrid   female    30 Concorde120   120      2.7
3     3 [17,18) Roland   male      30 Standard120   120      9.2
4     4 [17,18) Isabella female    48 Standard120   120      6  
5     5 [17,18) Fabienne female    17 Standard120   120      7.3
6     6 [17,18) Fabienne female    17 Standard120   120      7.8
```

Ensuring 'paper' is factor, not numeric

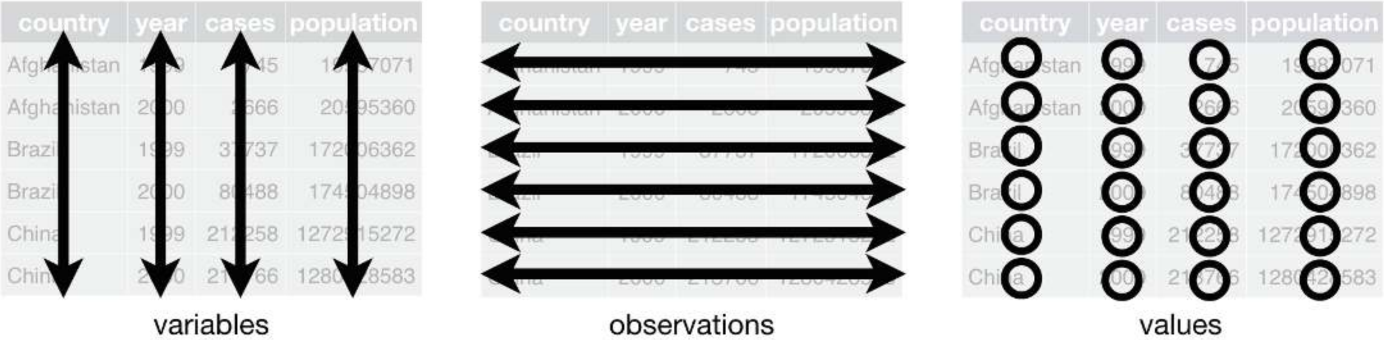
Using dplyr:

```
paperplanes <- paperplanes %>%
  mutate(paper = as.factor(paper))
```

R base:

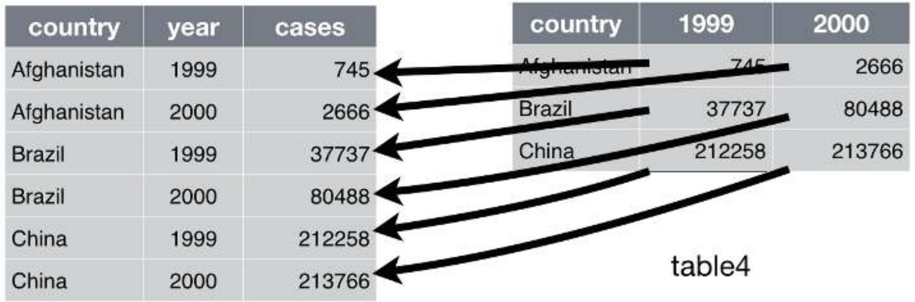
```
paperplanes$paper <- as.factor(paperplanes$paper)
```

Data must be a tidy data frame



country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

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



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

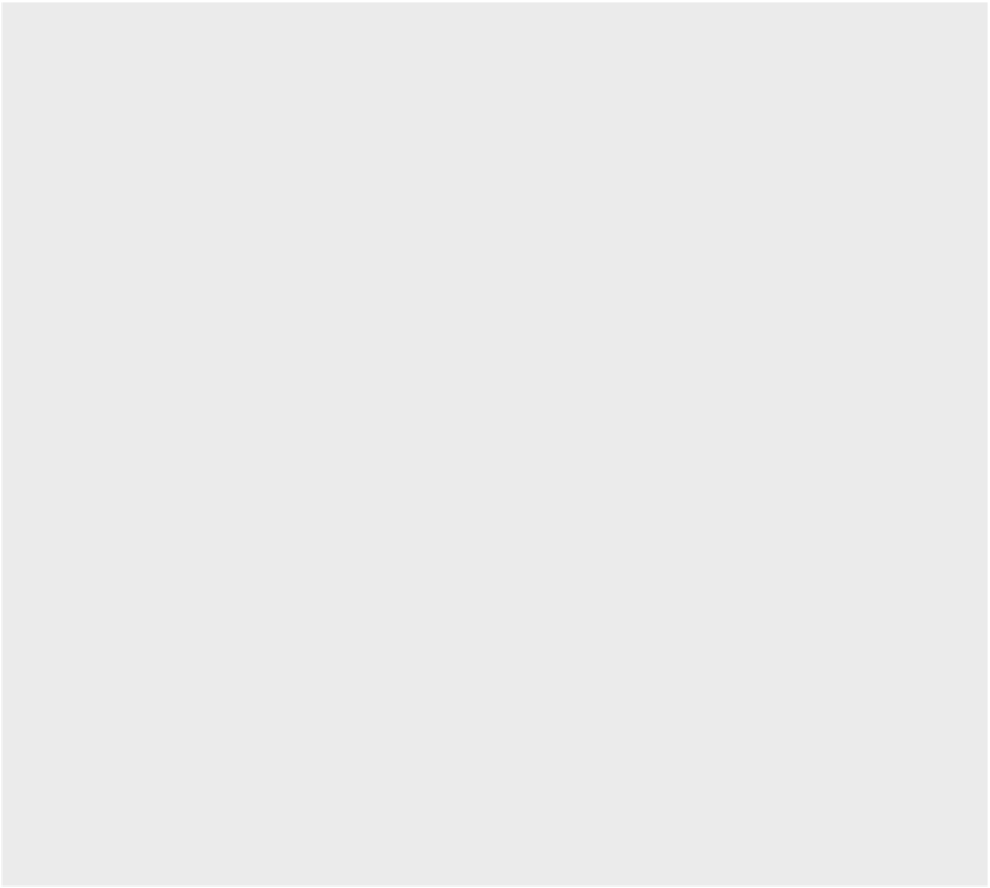
country	1999	2000
Afghanistan	745	2666
Brazil	37737	80488
China	212258	213766

table4

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

Calling ggplot

```r
library(ggplot2)
ggplot(paperplanes)
```



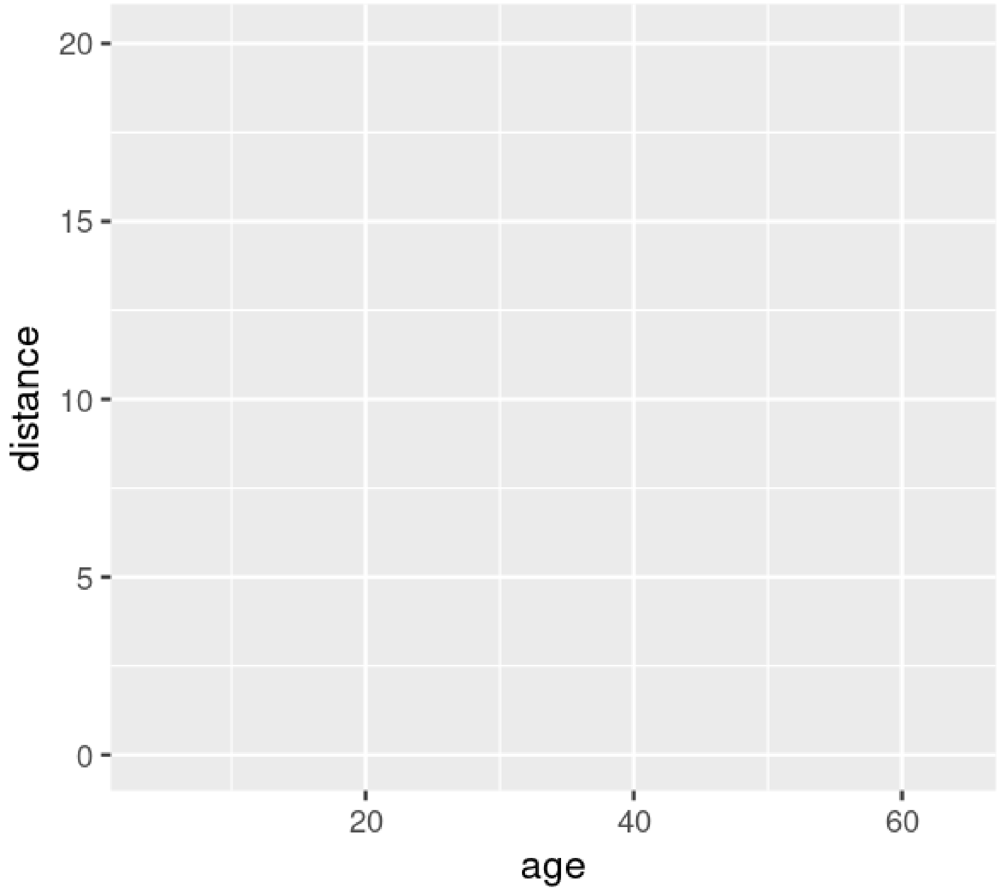
First argument is a tidy data frame

```
ggplot(paperplanes)
```

What variables as axes?

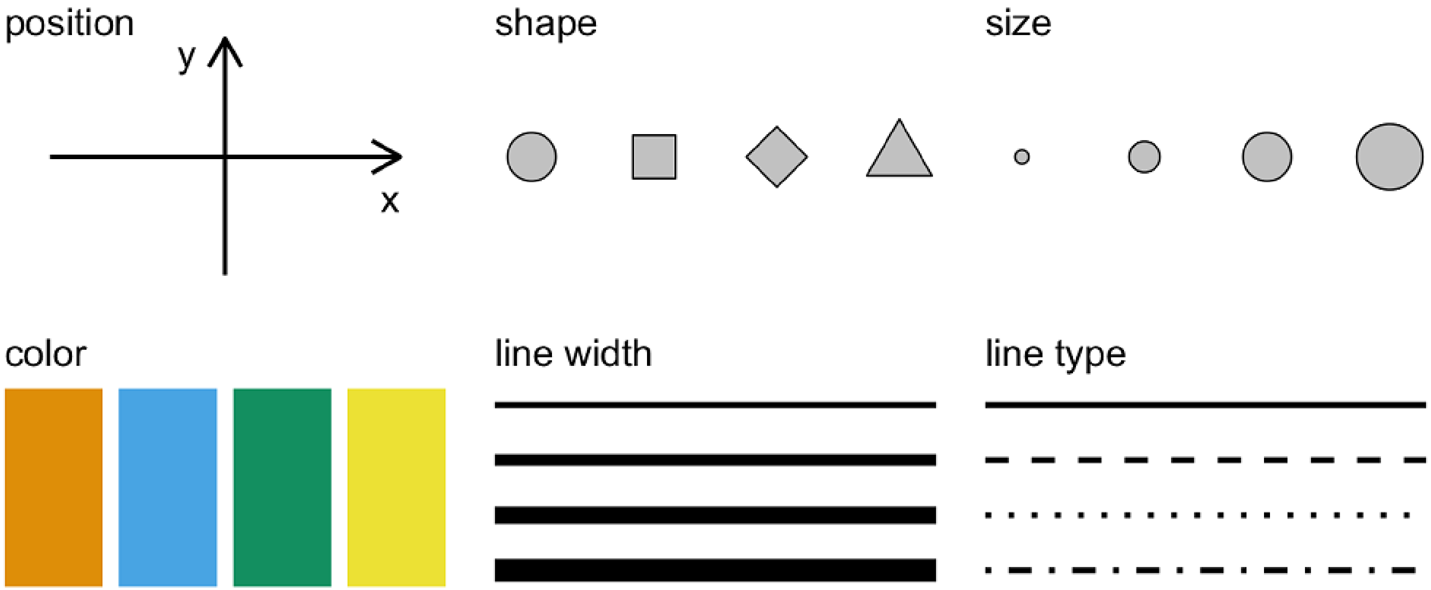
Note syntax: + followed by new line

```
ggplot(paperplanes) +
  aes(x = age, y = distance)
```



Aesthetics (*aes*) map data variables (*age, distance*) to graphic elements (*axes*)

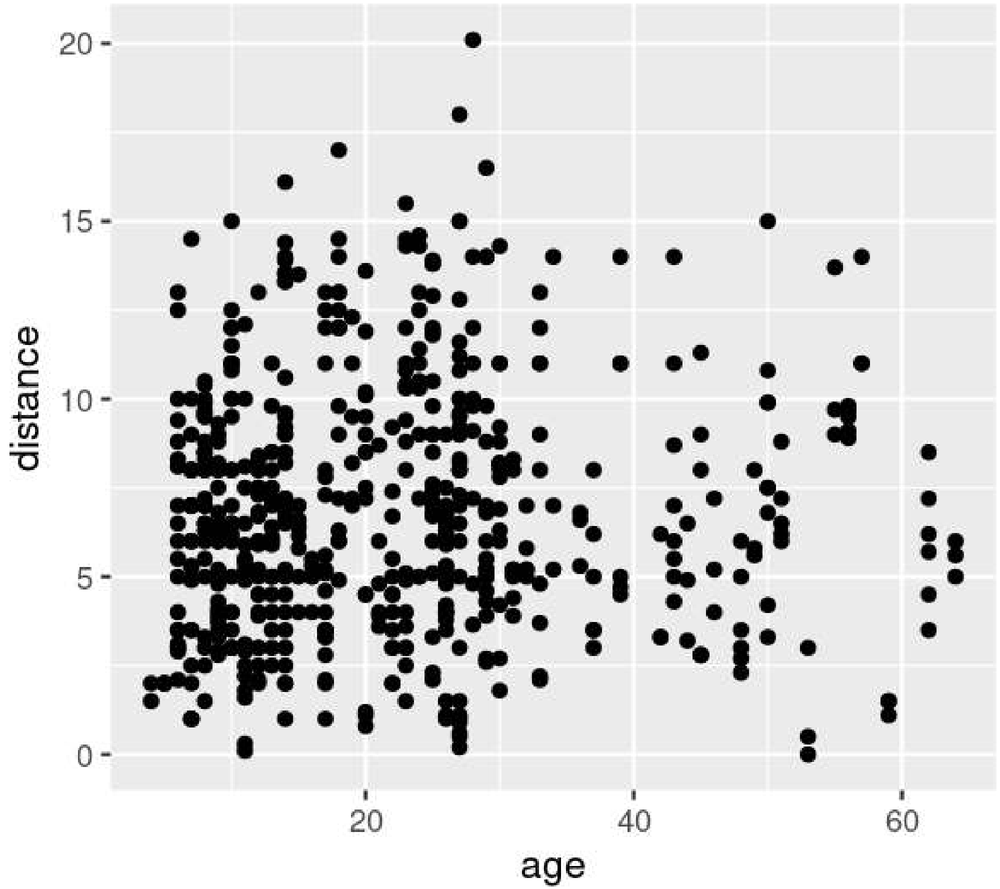
```
ggplot(paperplanes) +
  aes(x = age, y = distance)
```



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

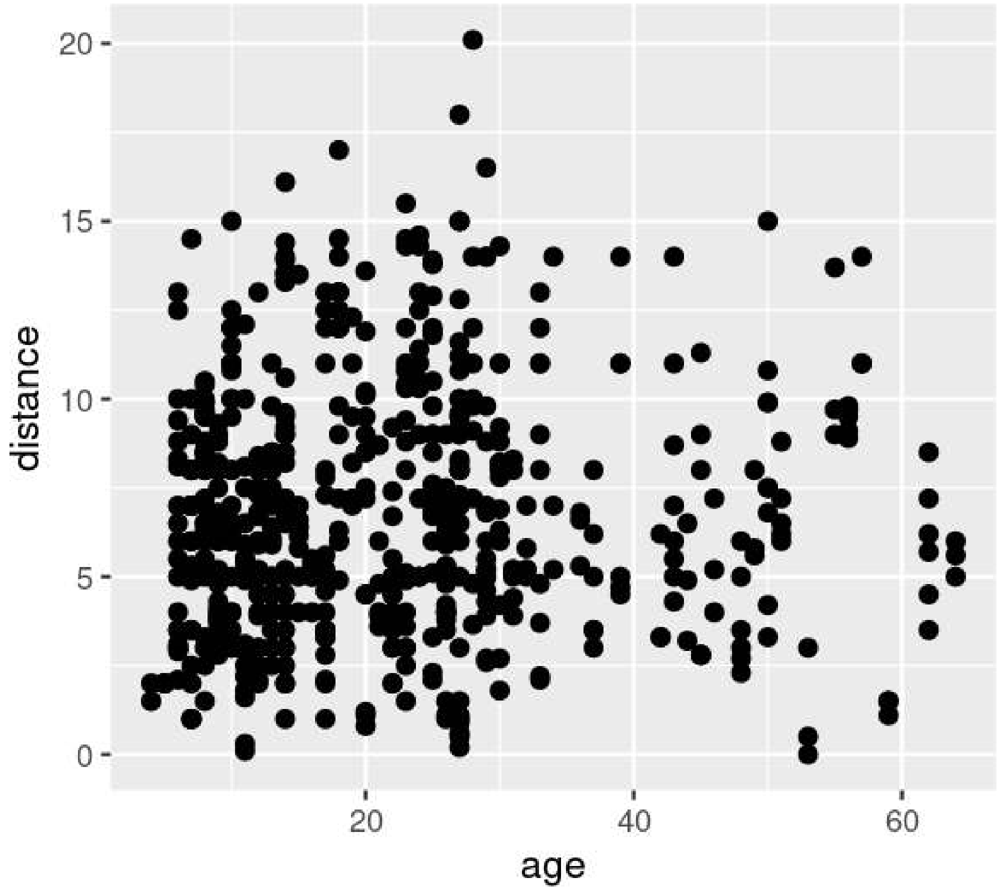
Add layers (geoms)

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point()
```



Change point size and type

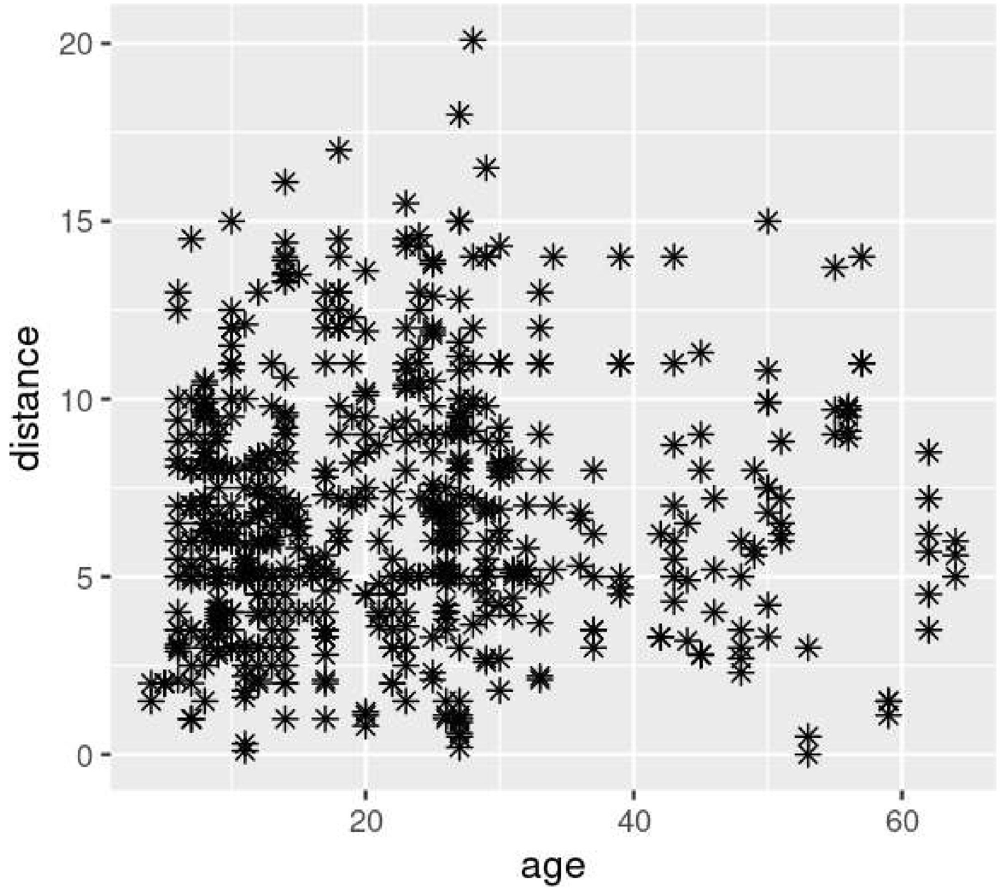
```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(size = 2)
```



Check out `geom_point` help here

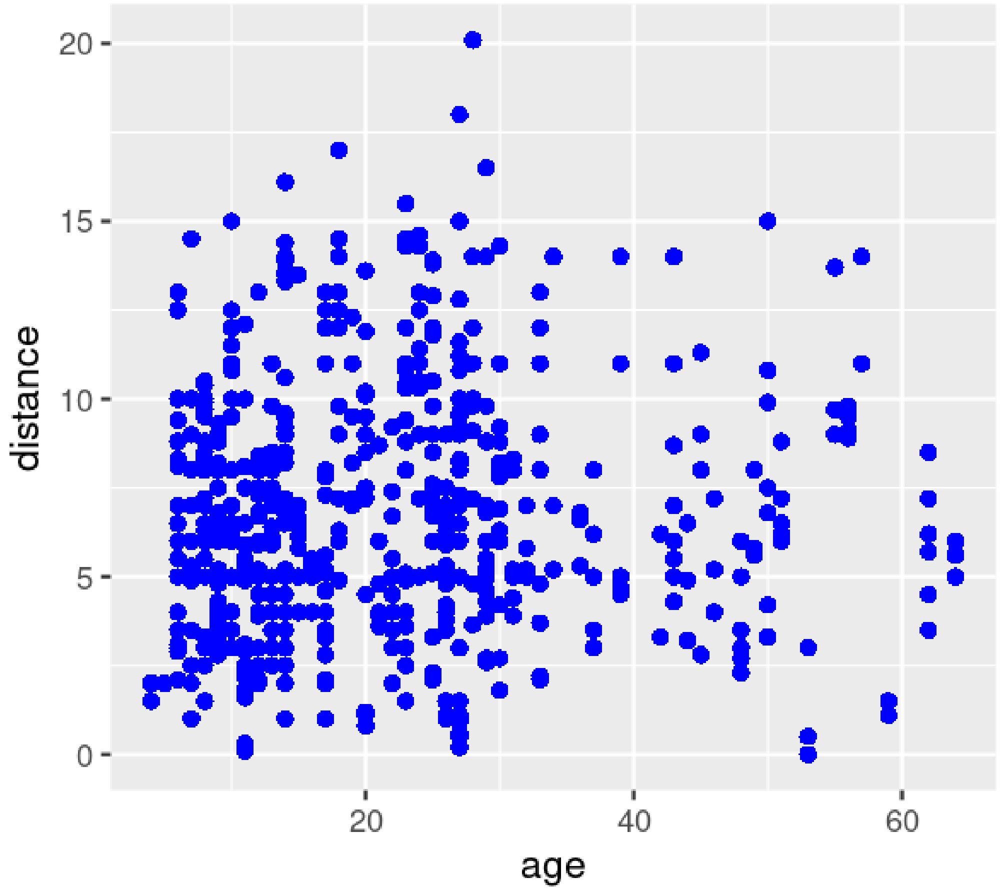
Change point size and type

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(size = 2, shape = 8)
```



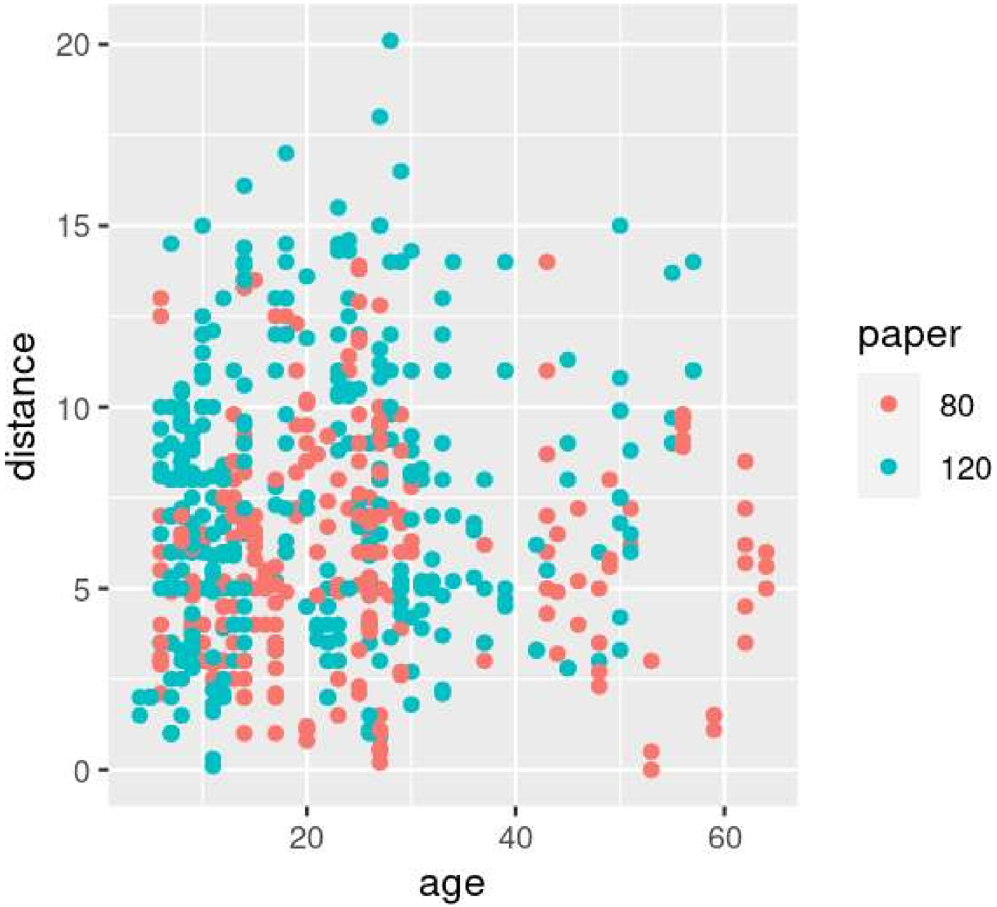
Change point size and type

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(size = 2, shape = 16,
                 colour = "blue")
```



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

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper))
```



Remember:

'aes' relates some graphical characteristic

(colour, size, shape...)

to a variable in the data

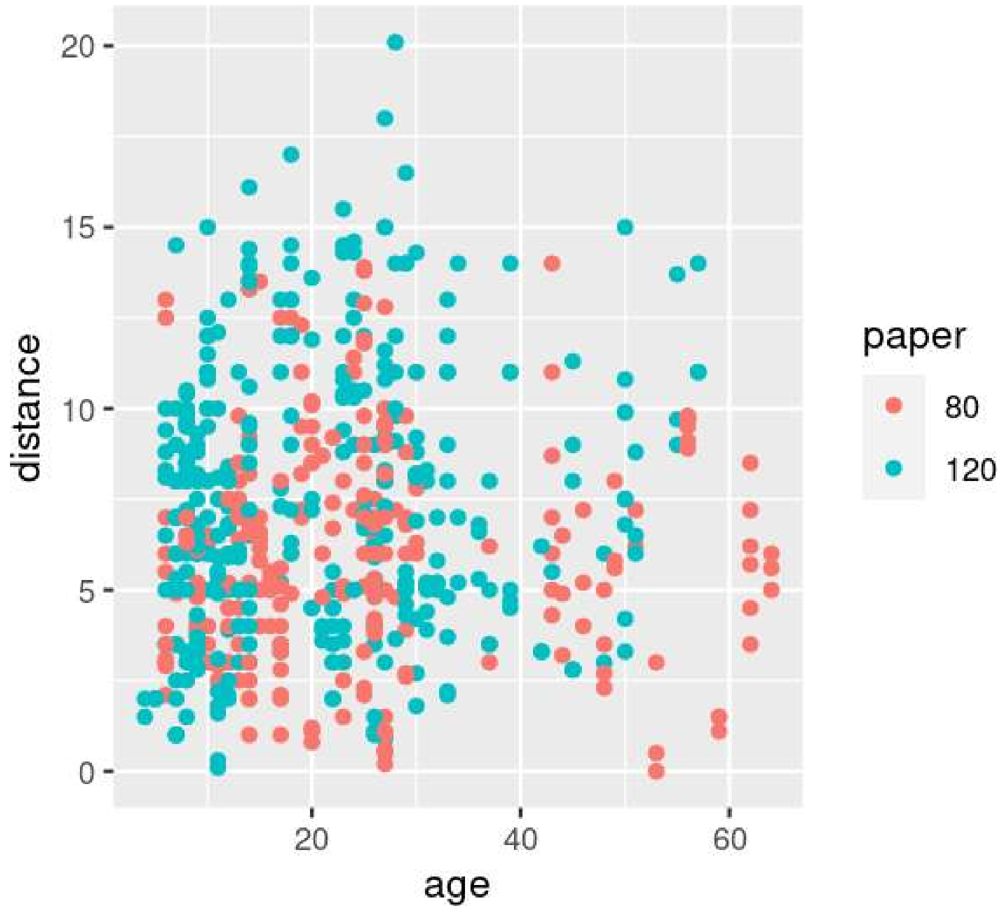
Note difference between

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

```
geom_point(aes(colour = gender))
# colour maps a *variable* in the data (gender) USING `aes`
```

This works:

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper))
```



This doesn't work:

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(colour = paper)
```

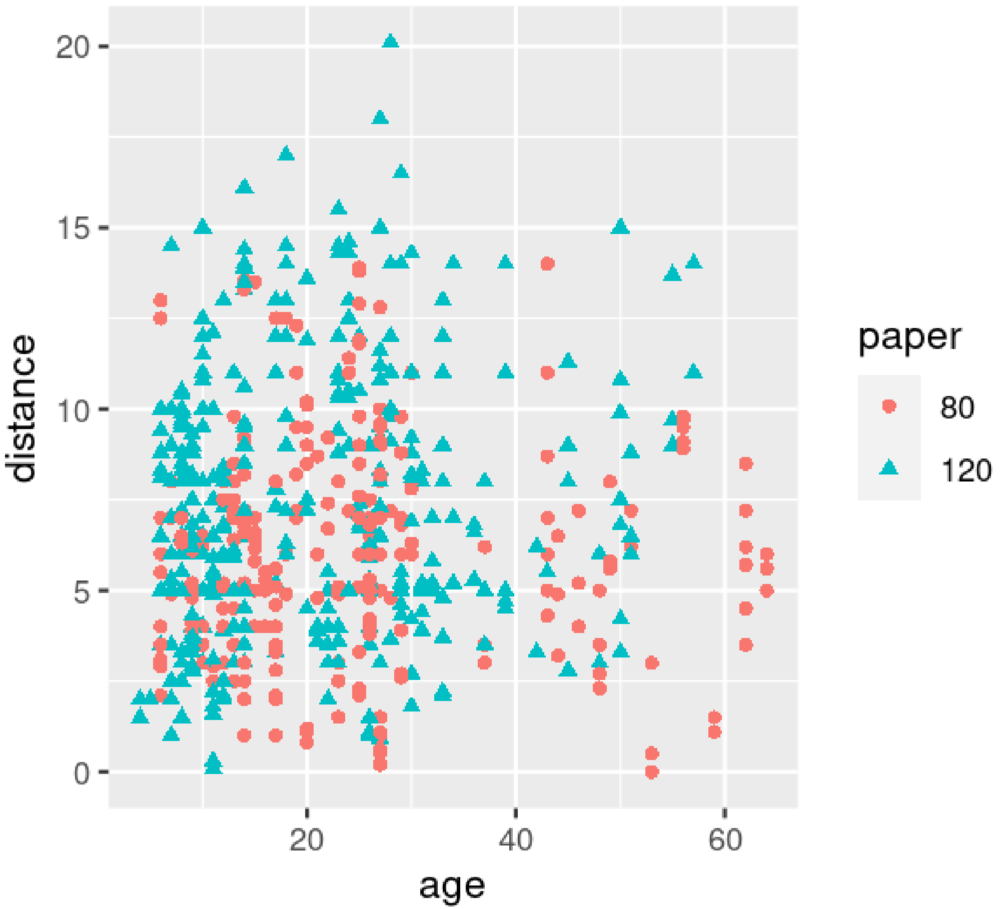
Error in layer(data = data, mapping = mapping, stat = stat, geom = GeomPoint, : object 'paper' not found

'paper' is a variable in dataframe, hence

must use `aes`

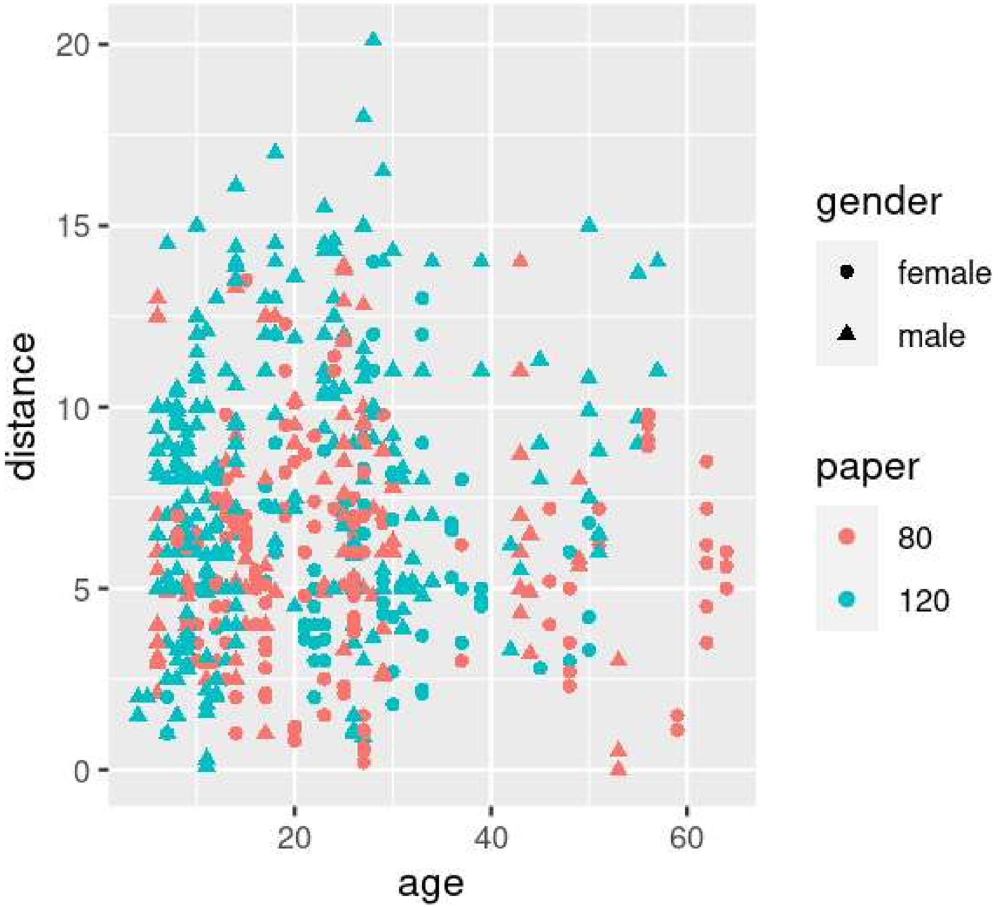
Map geom aesthetics (colour, shape) to variable

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper,
                 shape = paper))
```



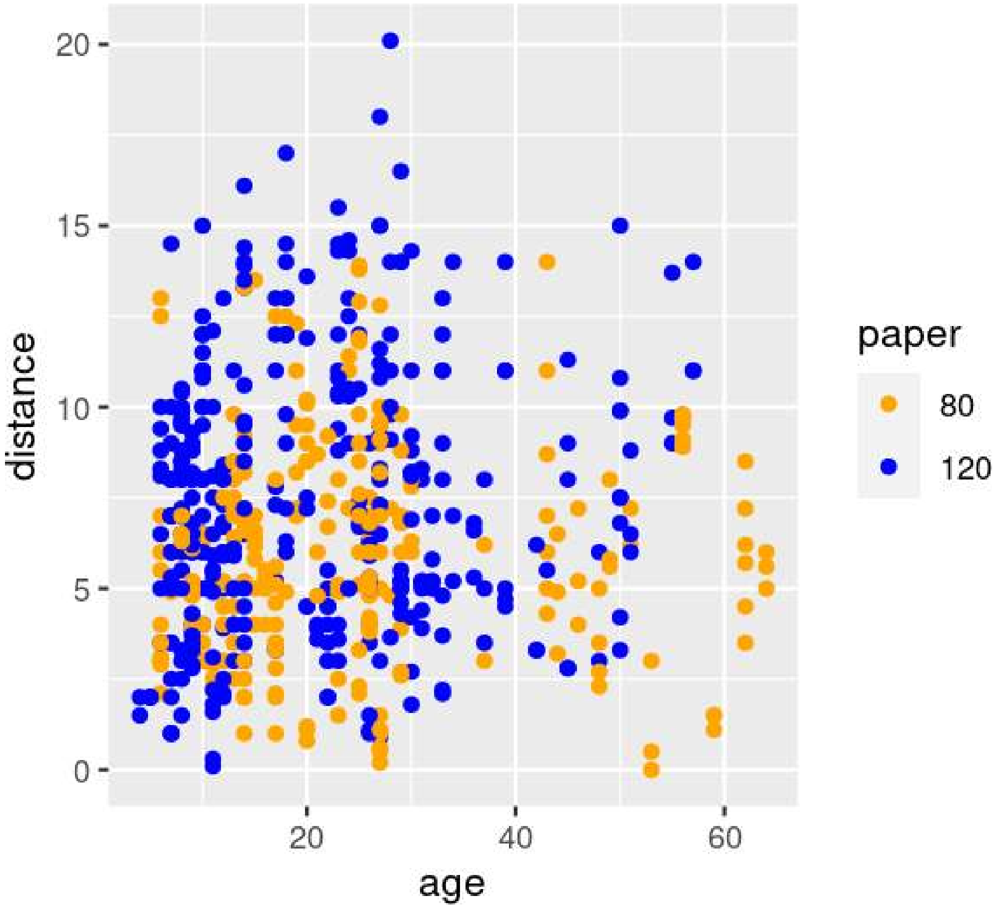
Map geom aesthetics (colour, shape) to variable

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper,
                 shape = gender))
```



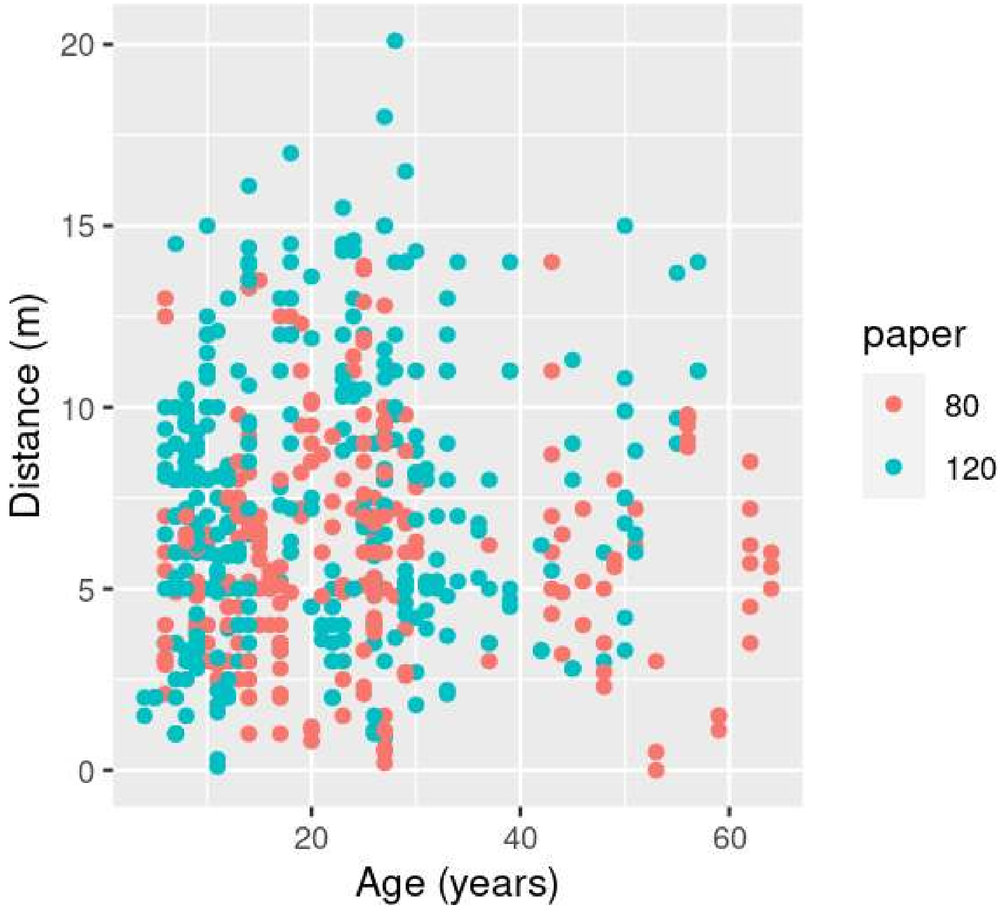
Change colour scale

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  scale_colour_manual(values = c("orange", "b
```



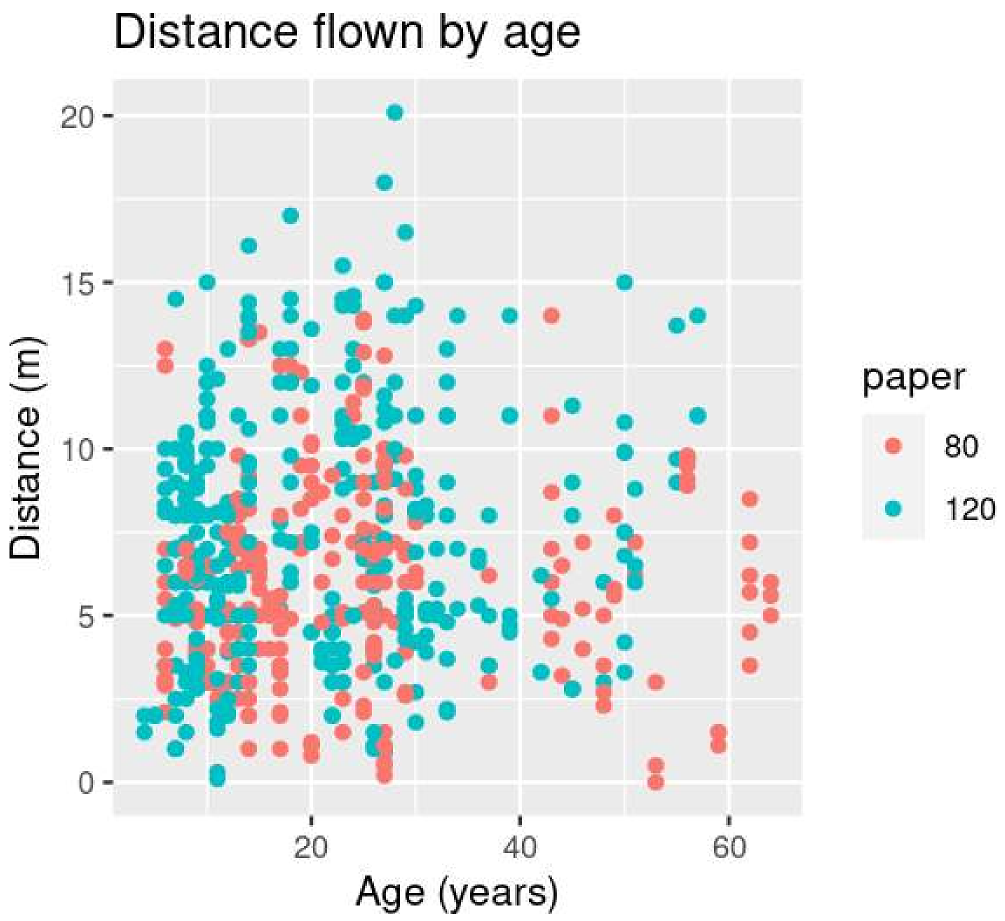
Change axis labels: xlab & ylab

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  labs(x = "Age (years)",
       y = "Distance (m)")
```



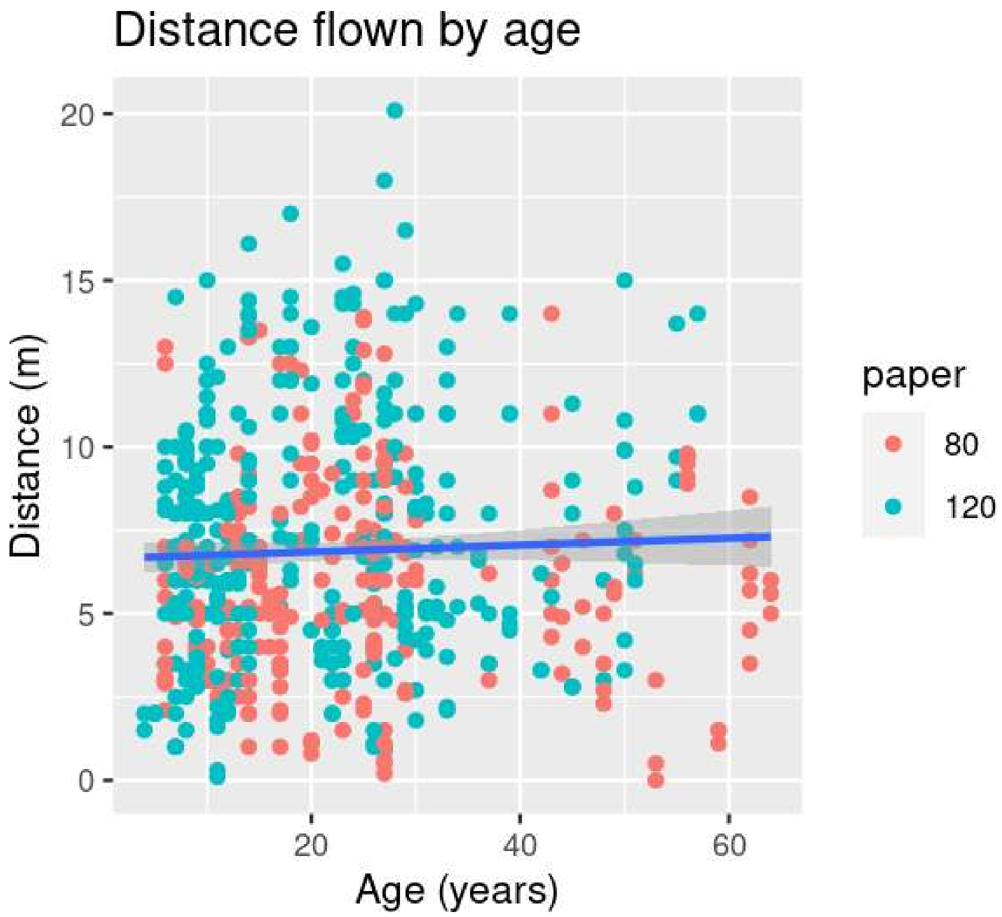
Set title

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  labs(x = "Age (years)",
       y = "Distance (m)") +
  labs(title = "Distance flown by age")
```



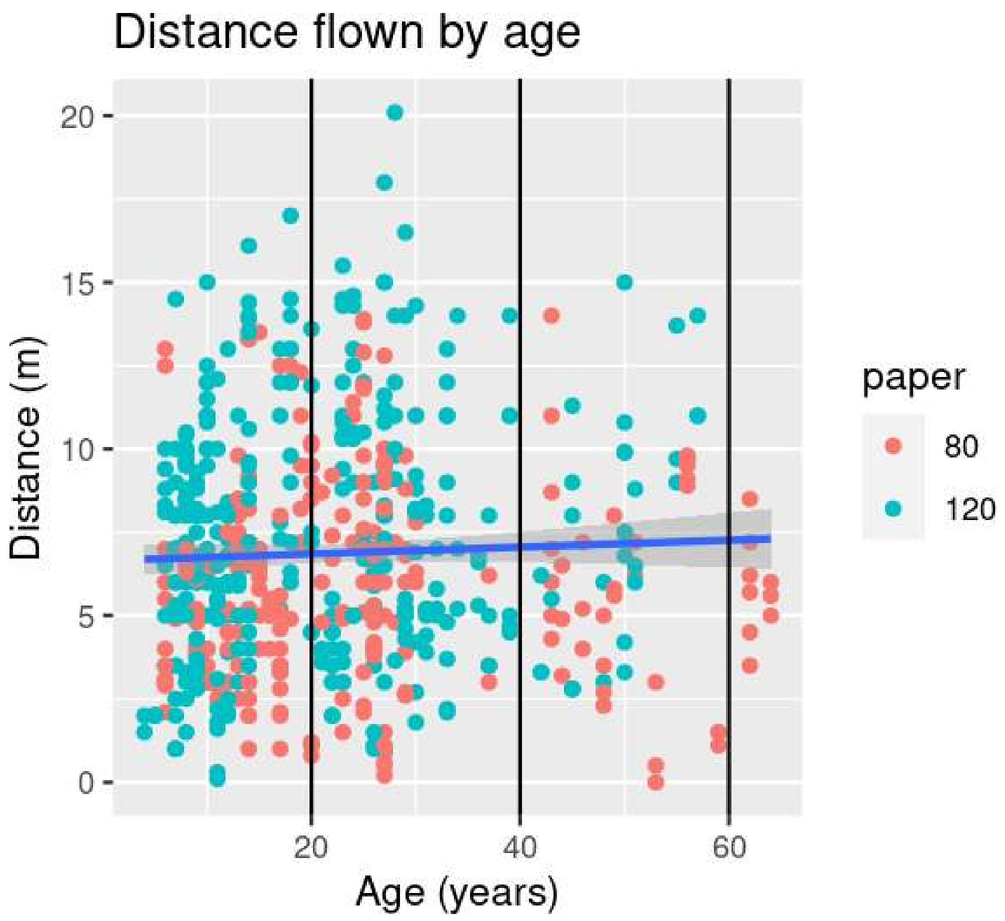
Adding more layers

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  labs(x = "Age (years)",
       y = "Distance (m)") +
  labs(title = "Distance flown by age") +
  geom_smooth(method = "lm")
```



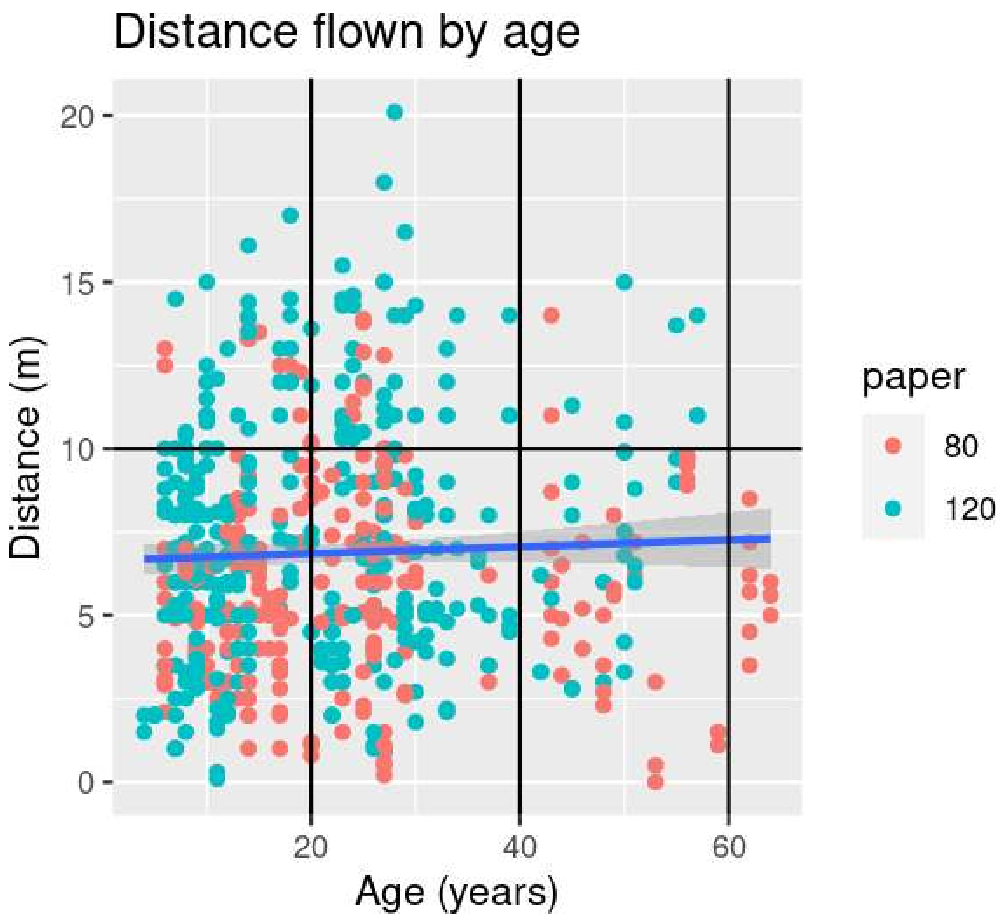
Adding more layers

```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  labs(x = "Age (years)",
       y = "Distance (m)") +
  labs(title = "Distance flown by age") +
  geom_smooth(method = "lm") +
  geom_vline(xintercept = c(20, 40, 60))
```



Adding more layers

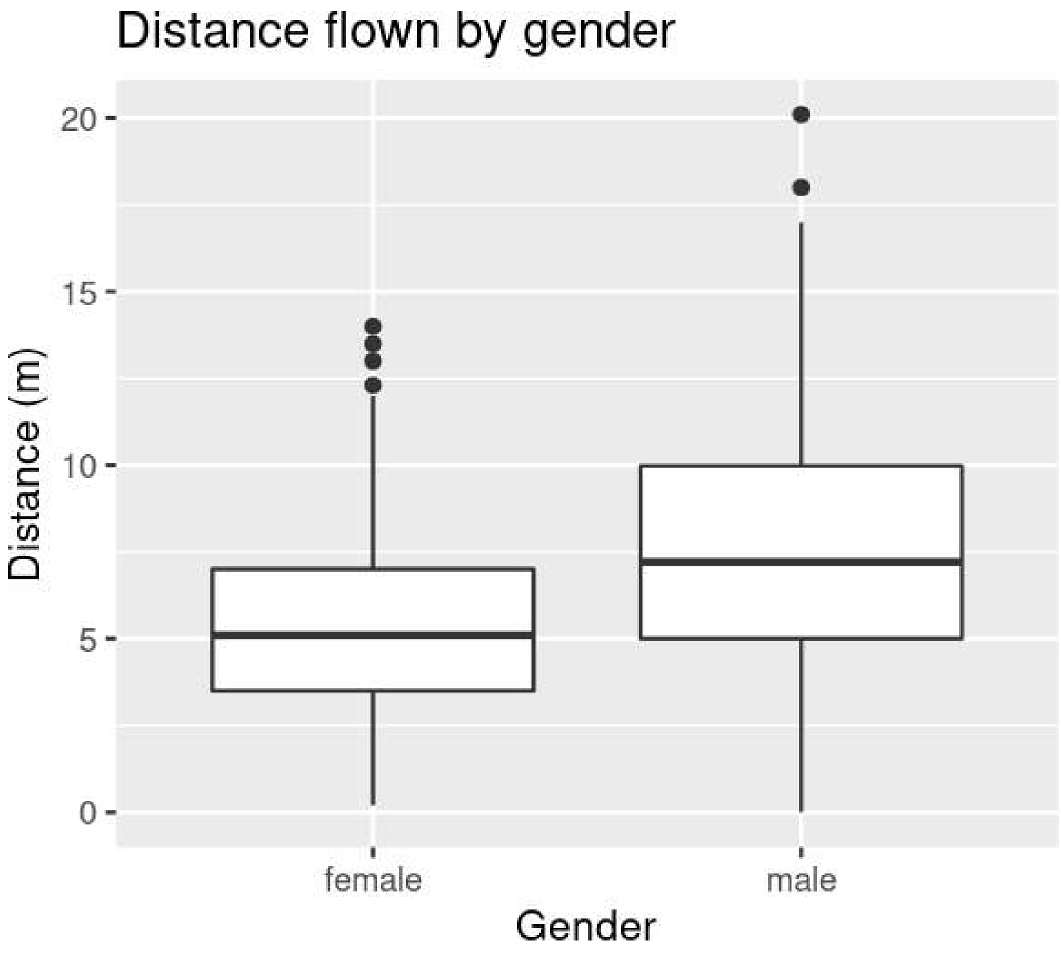
```
ggplot(paperplanes) +
  aes(x = age, y = distance) +
  geom_point(aes(colour = paper)) +
  labs(x = "Age (years)",
       y = "Distance (m)") +
  labs(title = "Distance flown by age") +
  geom_smooth(method = "lm") +
  geom_vline(xintercept = c(20, 40, 60)) +
  geom_hline(yintercept = 10)
```



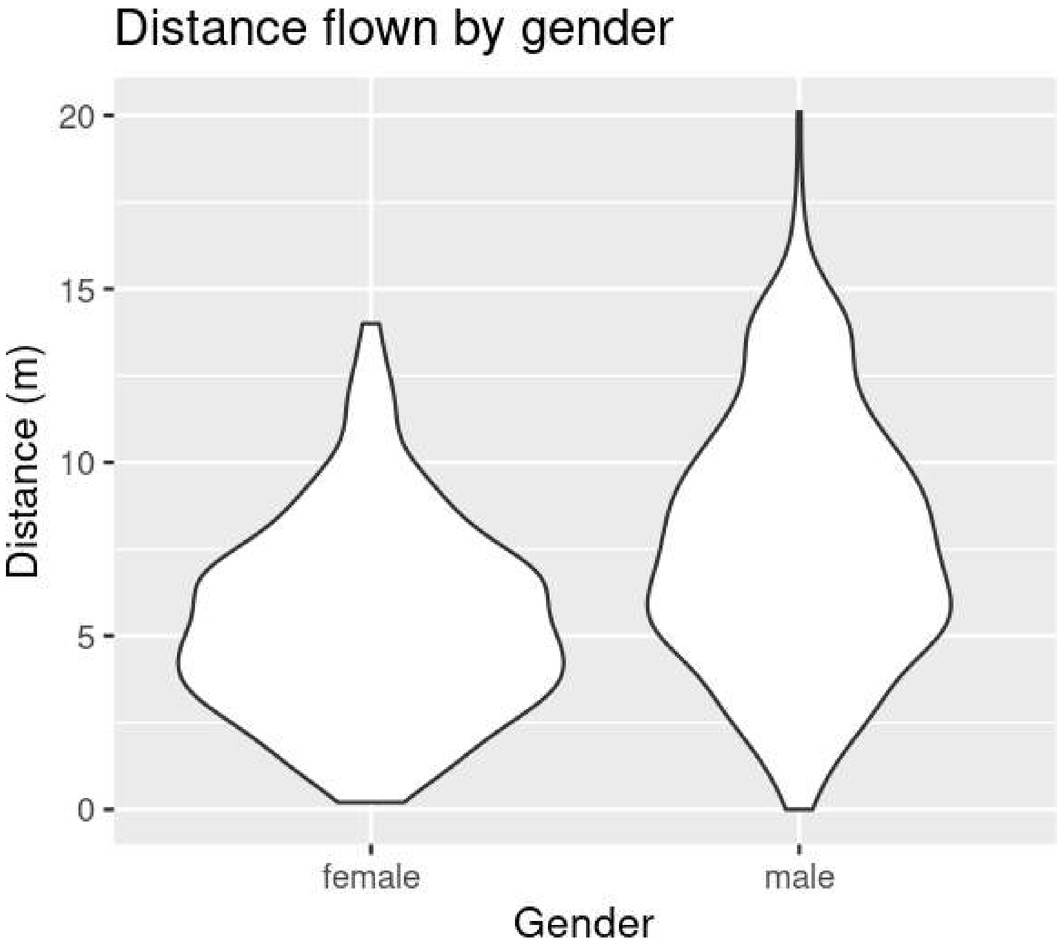
Summary

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

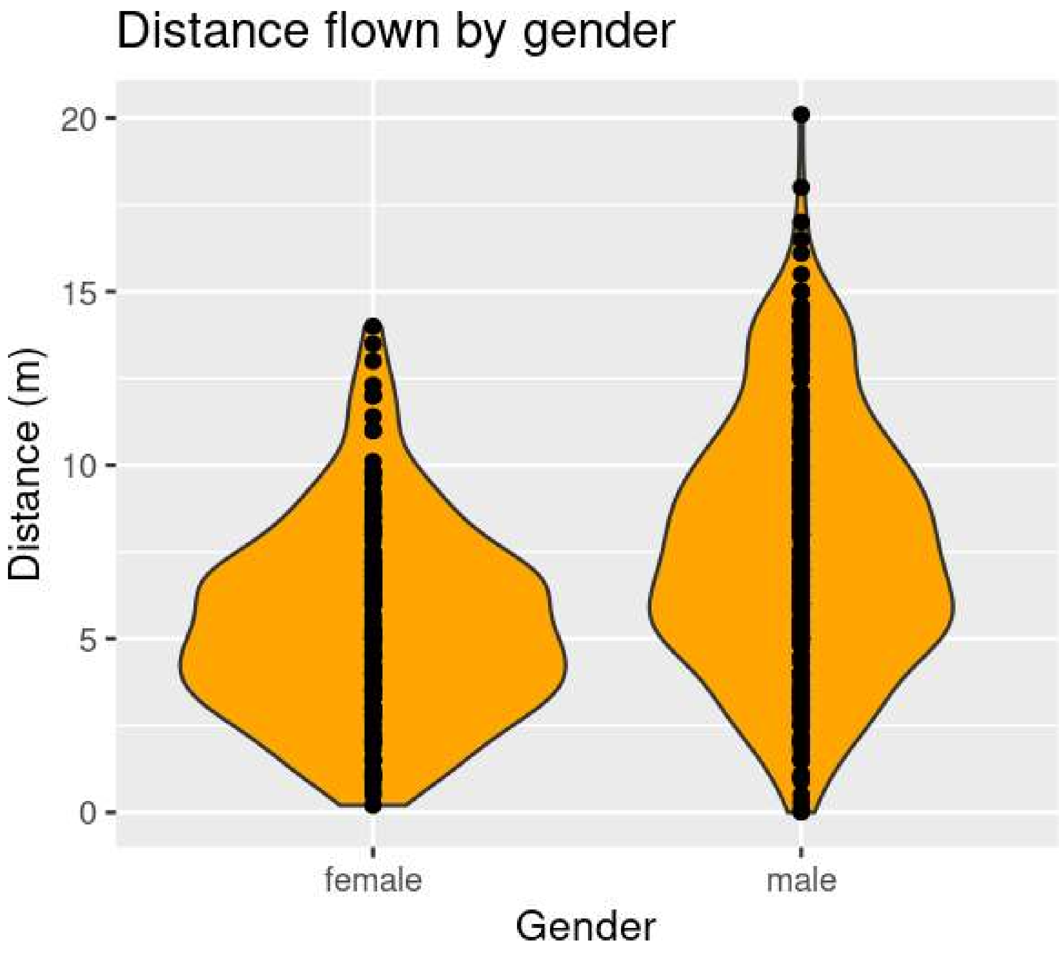
Exercise: Make a plot like this one



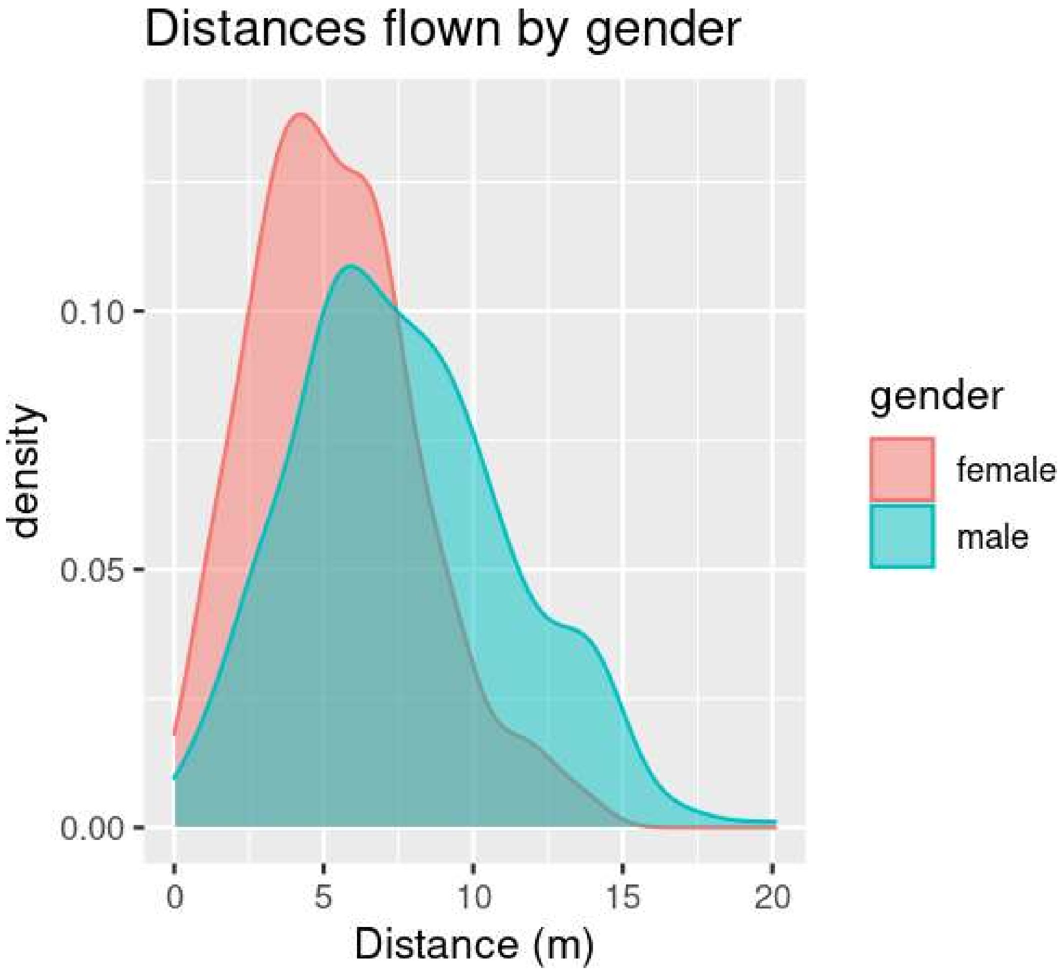
Exercise: Make a plot like this one



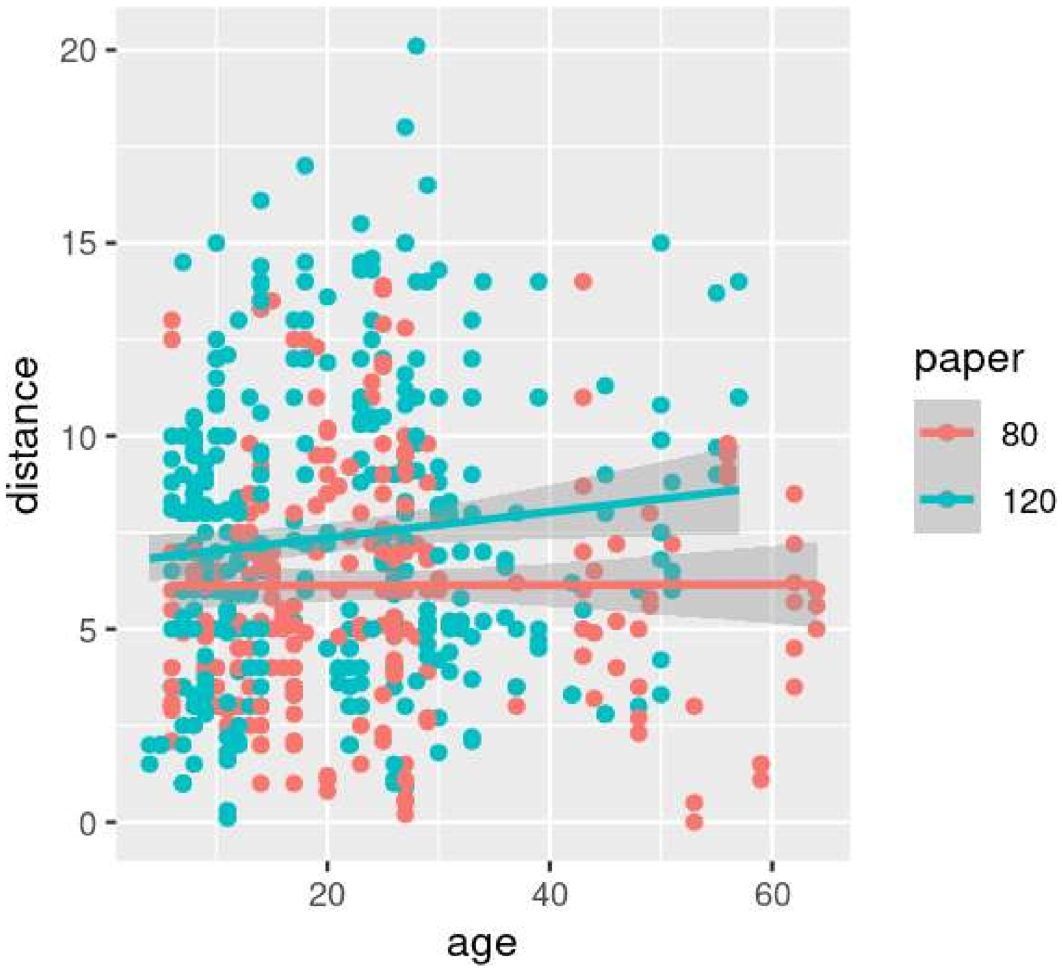
Exercise: Make a plot like this one



Exercise: Make a plot like this one



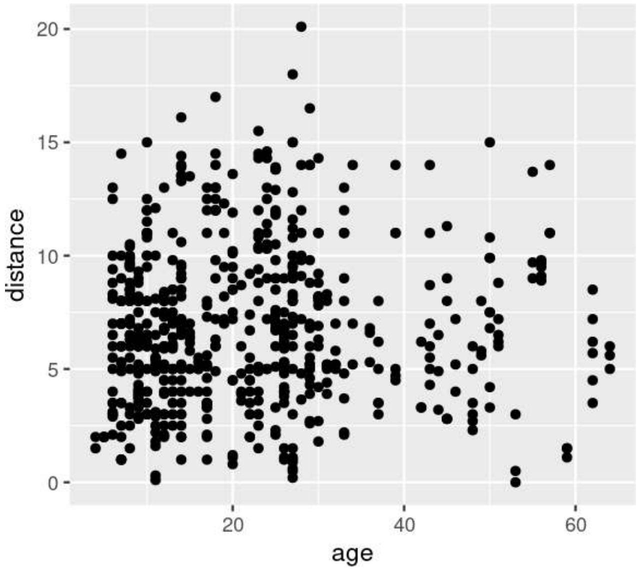
Exercise: Make a plot like this one



ggplot2 figures can be assigned to R objects

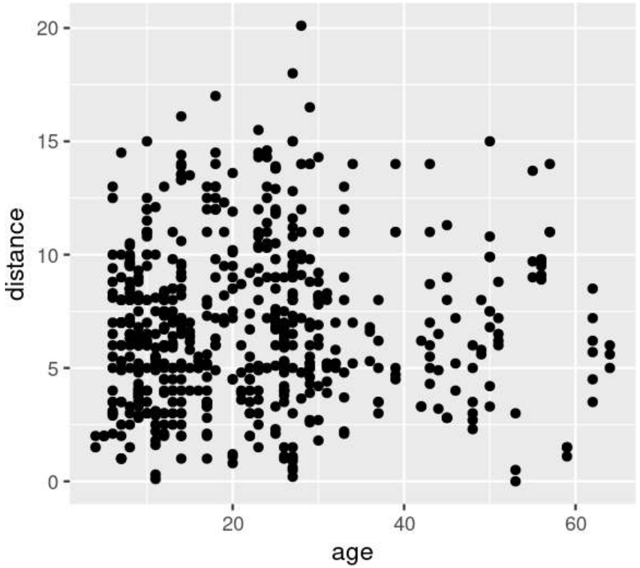
Assigning ggplot objects

```
myplot <- ggplot(paperplanes) +
  aes(x = age, y = distance)
myplot + geom_point()
```



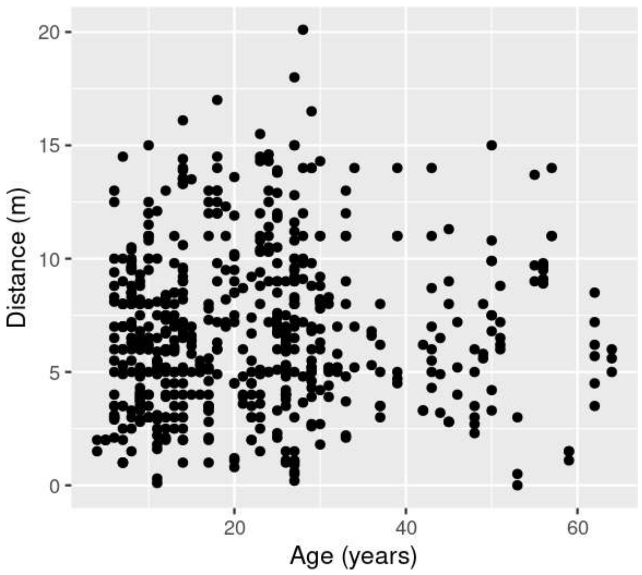
Assigning ggplot objects

```
myplot <- ggplot(paperplanes) +
  aes(x = age, y = distance)
myplot <- myplot + geom_point()
myplot
```



Assigning ggplot objects

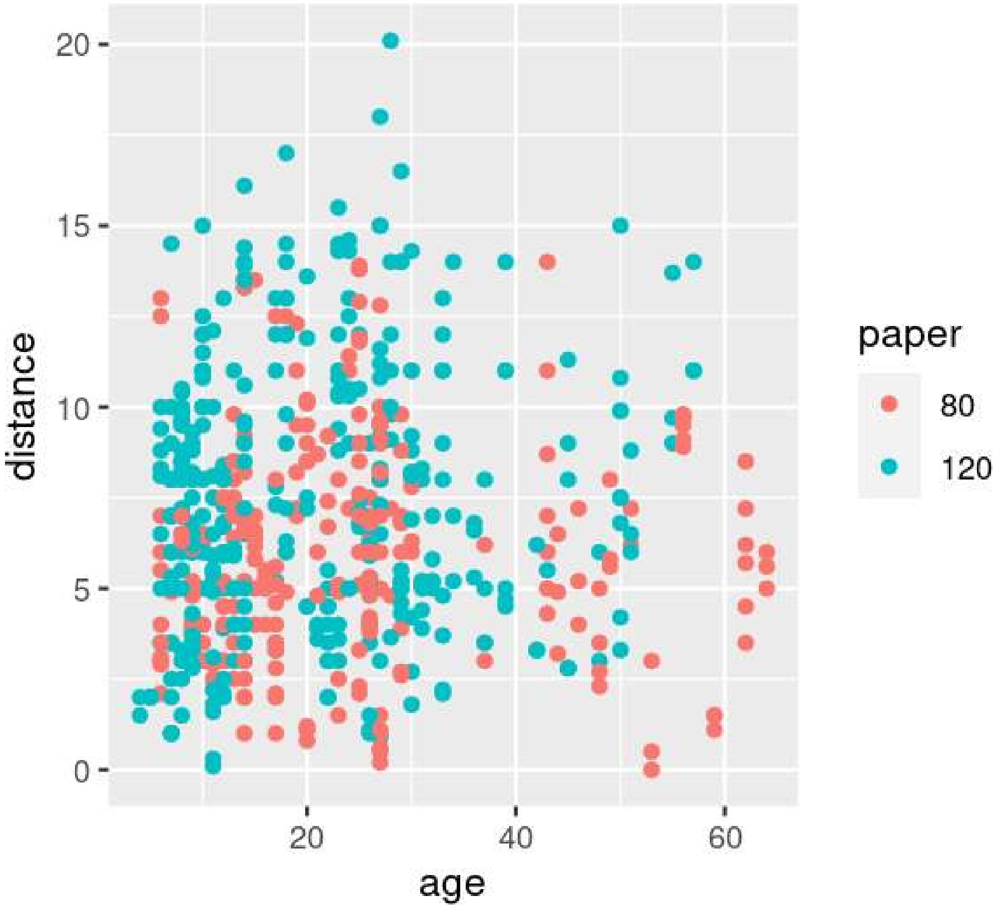
```
baseplot <- ggplot(paperplanes) +
  aes(x = age, y = distance)
scatterplot <- baseplot + geom_point()
labelled <- scatterplot + labs(x = "Age (years)", y = "Distance (m)")
labelled
```



Themes: changing plot appearance

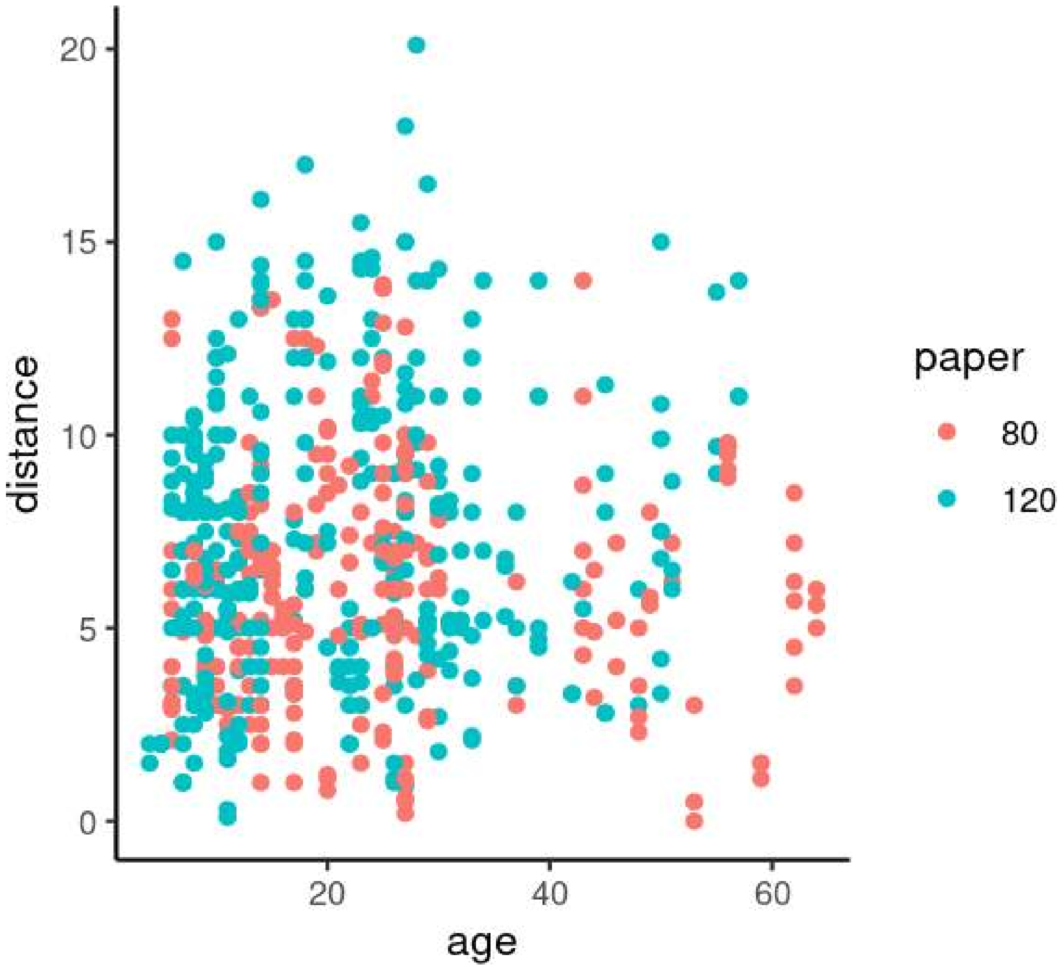
Create 'myplot'

```
myplot <- ggplot(paperplanes) +
  aes(x = age,
      y = distance,
      colour = paper) +
  geom_point()

myplot
```



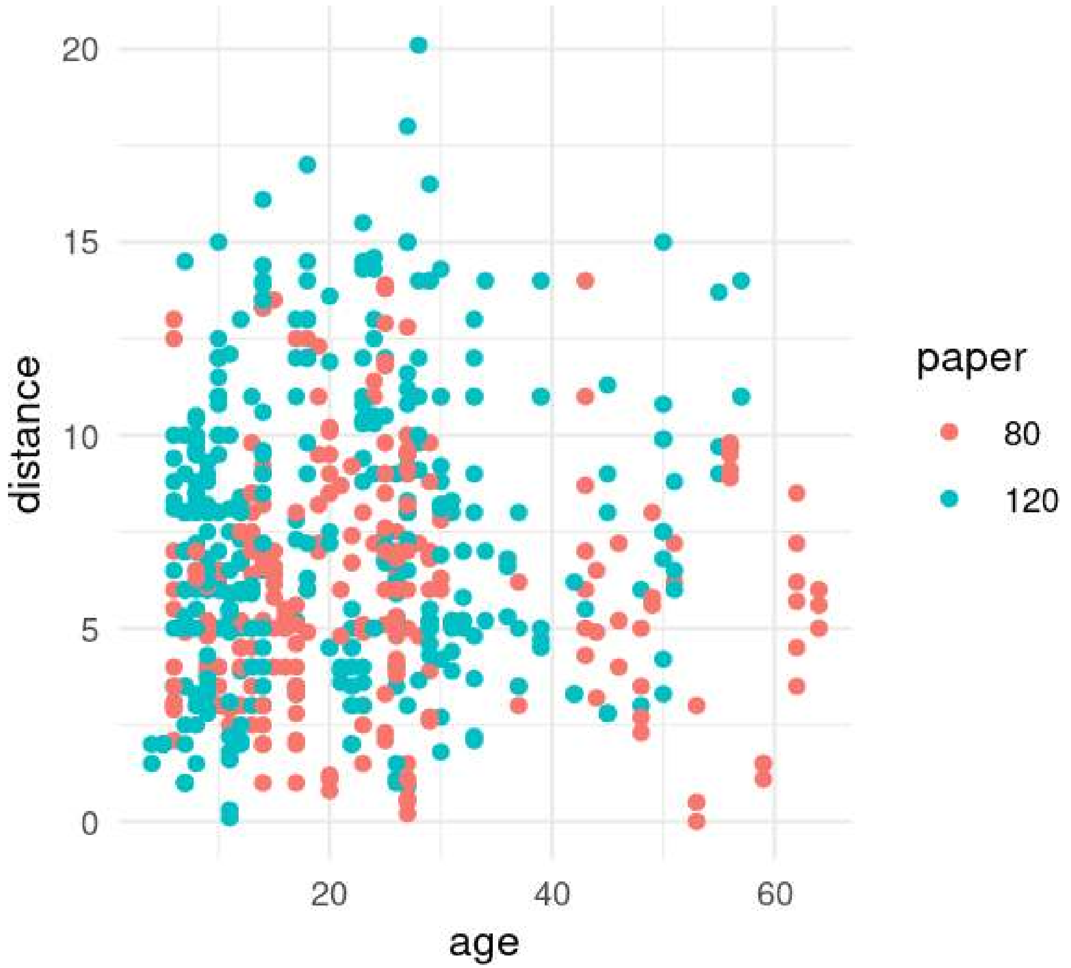
Use theme_classic

```
myplot + theme_classic()
```



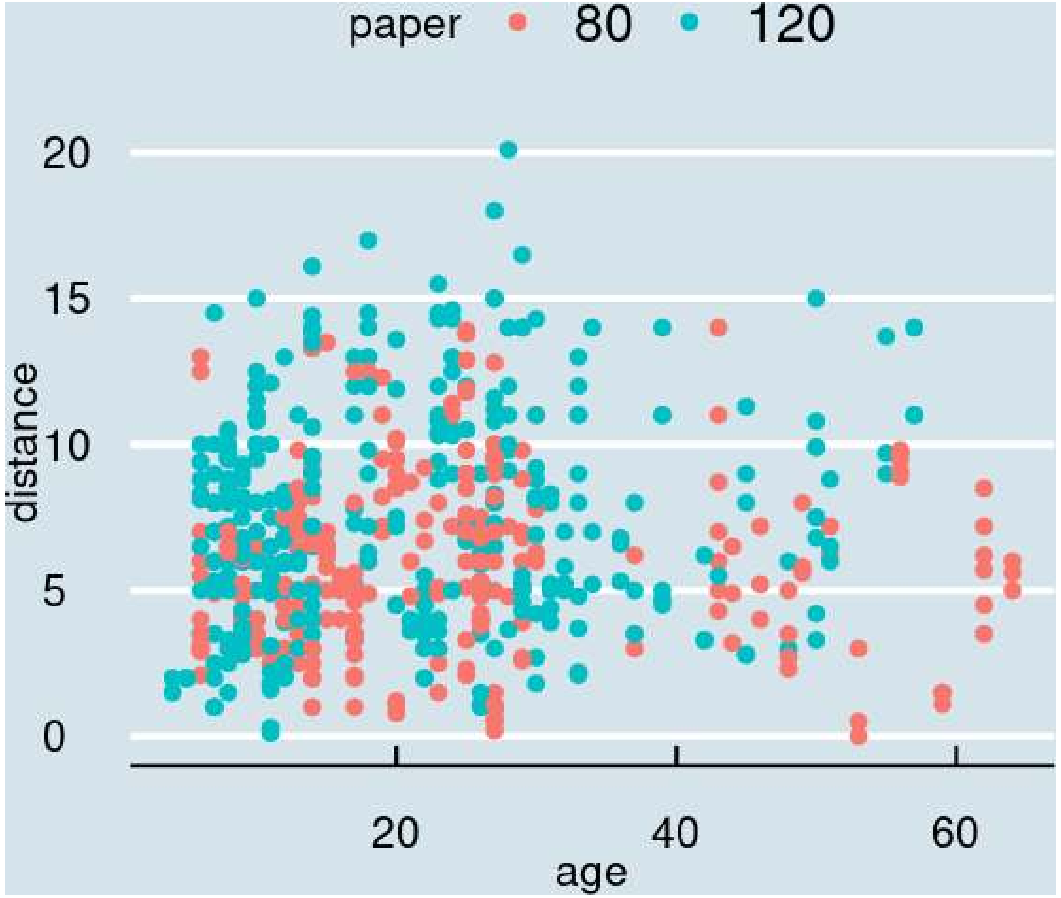
theme_minimal

```
myplot + theme_minimal()
```



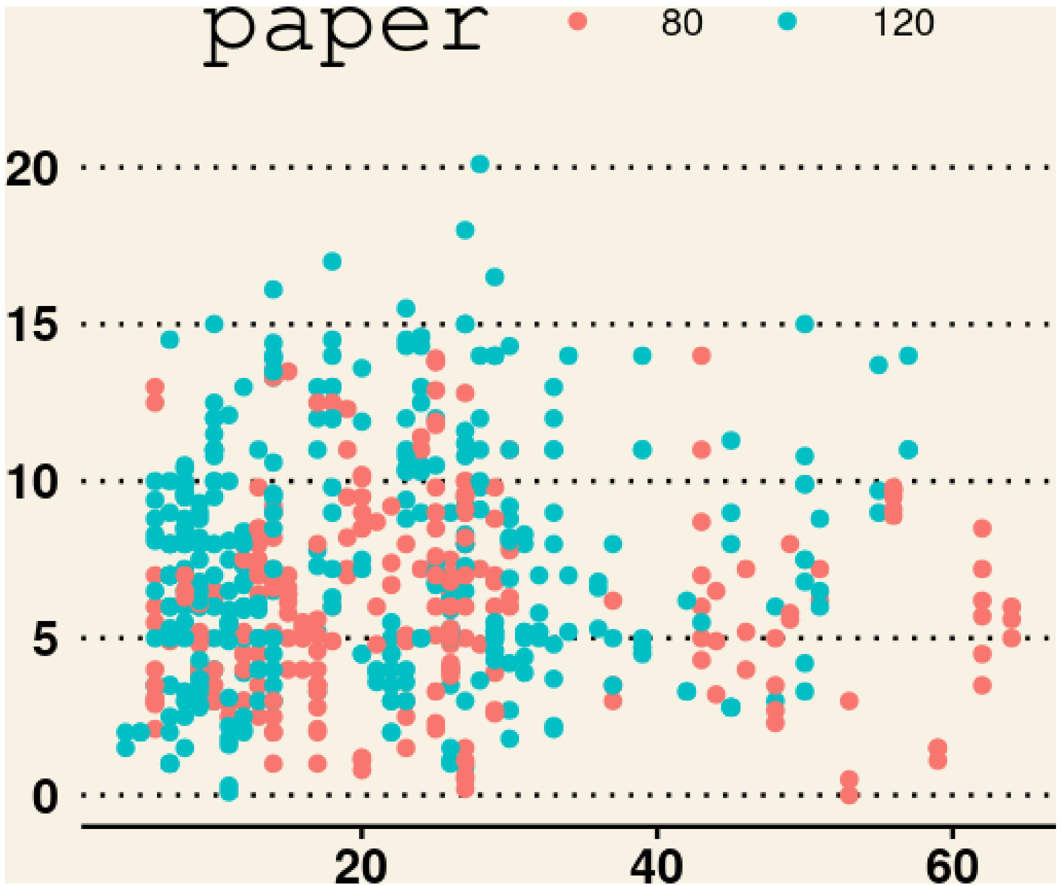
Lots of themes out there

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



Lots of themes out there

```
myplot + theme_wsj()
```

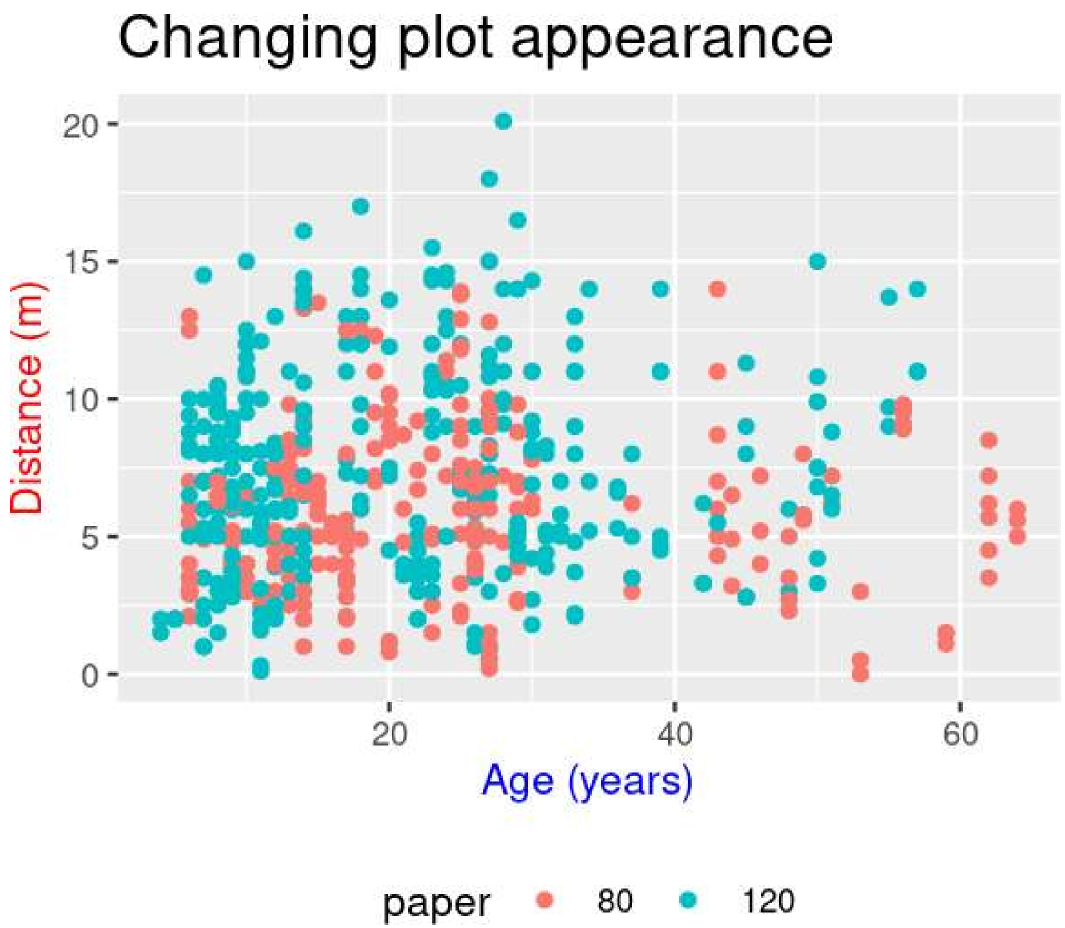


Editing themes

```
?theme
```

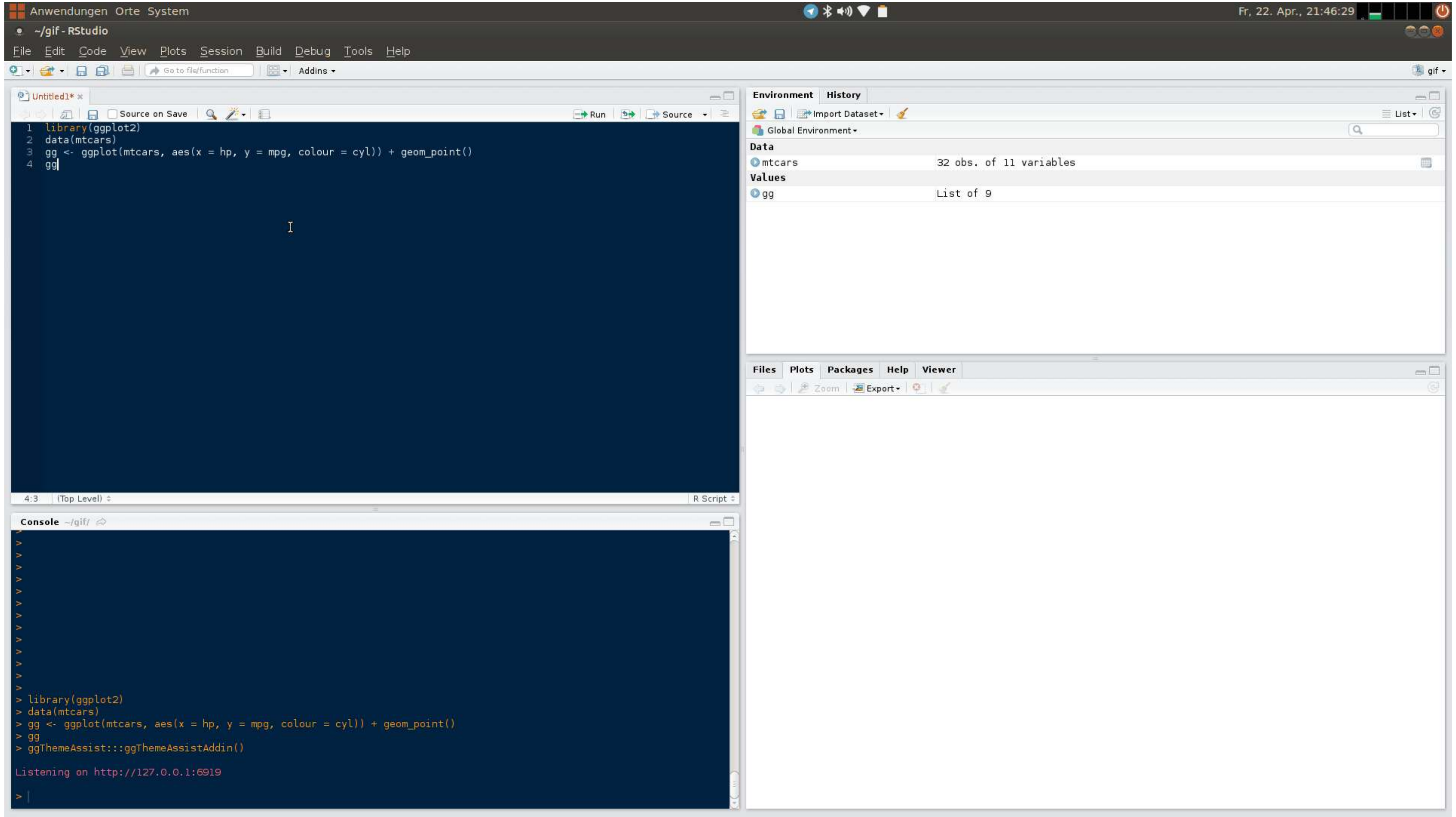
- `element_blank`
- `element_text`
- `element_line`
- `element_rect` (borders & backgrounds)

Exercise: make a plot like this one



Easily changing appearance with ggthemeassist (Rstudio addin)

https://github.com/calligross/ggthemeassist



```r
library(ggplot2)
data(mtcars)
gg <- ggplot(mtcars, aes(x = hp, y = mpg, colour = cyl)) + geom_point()
gg
```

Data

mtcars 32 obs. of 11 variables

Values

gg List of 9

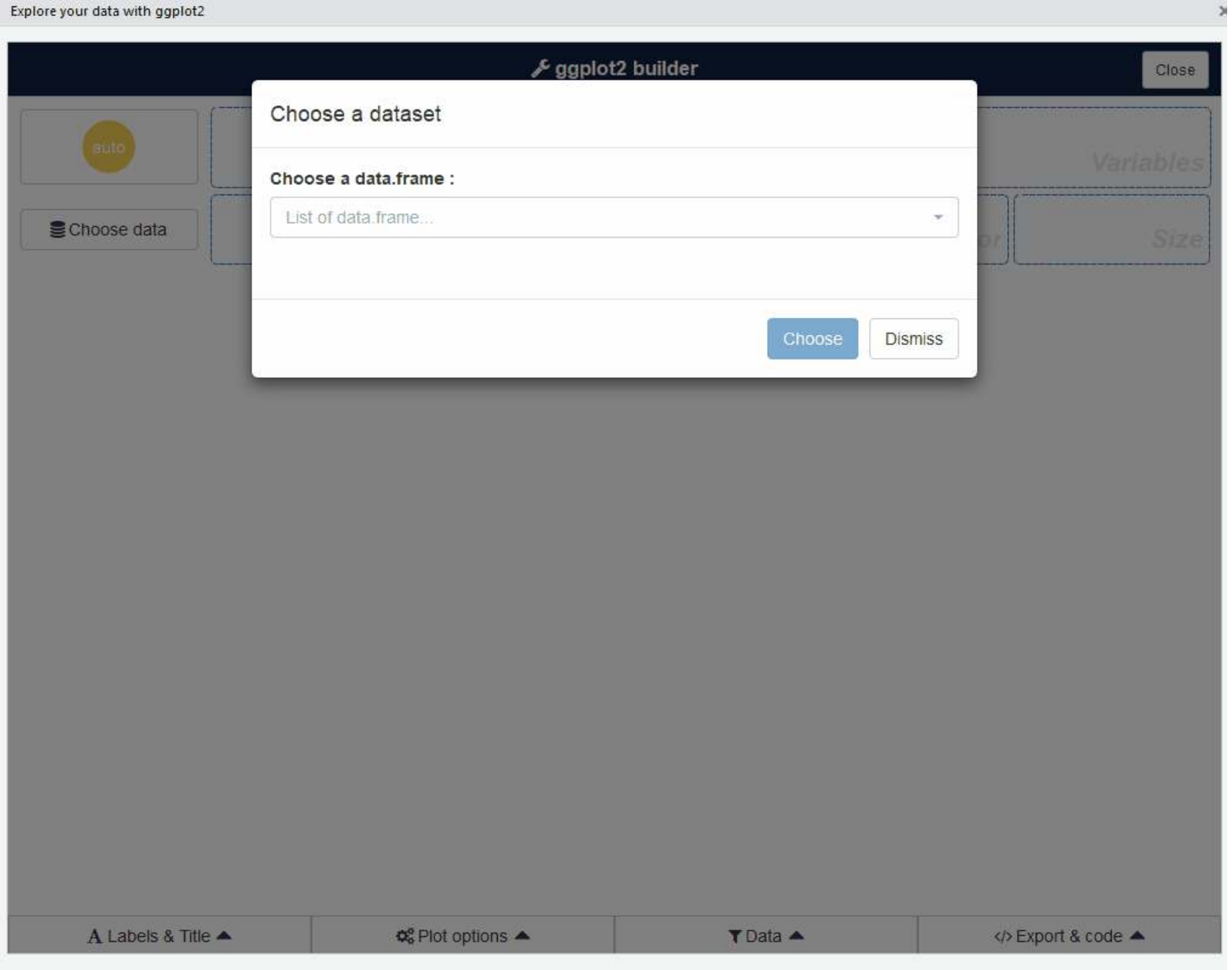
```
>
>
>
>
>
>
>
>
>
>
>
>
>
> library(ggplot2)
> data(mtcars)
> gg <- ggplot(mtcars, aes(x = hp, y = mpg, colour = cyl)) + geom_point()
> gg
> ggThemeAssist:::ggThemeAssistAddin()

Listening on http://127.0.0.1:6919

>
```

Easily changing appearance with ggedit

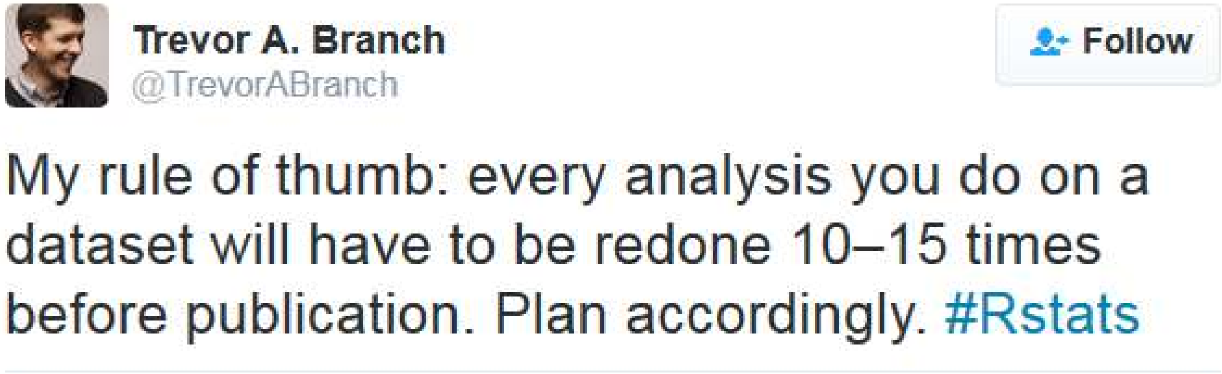
https://github.com/yonicd/ggedit

esquisse: ggplot2 builder addin

https://github.com/dreamRs/esquisse



Think twice before editing plots out of R



Trevor A. Branch
@TrevorABranch

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

Why I think twice before editing plots out of R

Choosing the right visualization software

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

```r
library(cowplot)
myplot + theme_cowplot()
```

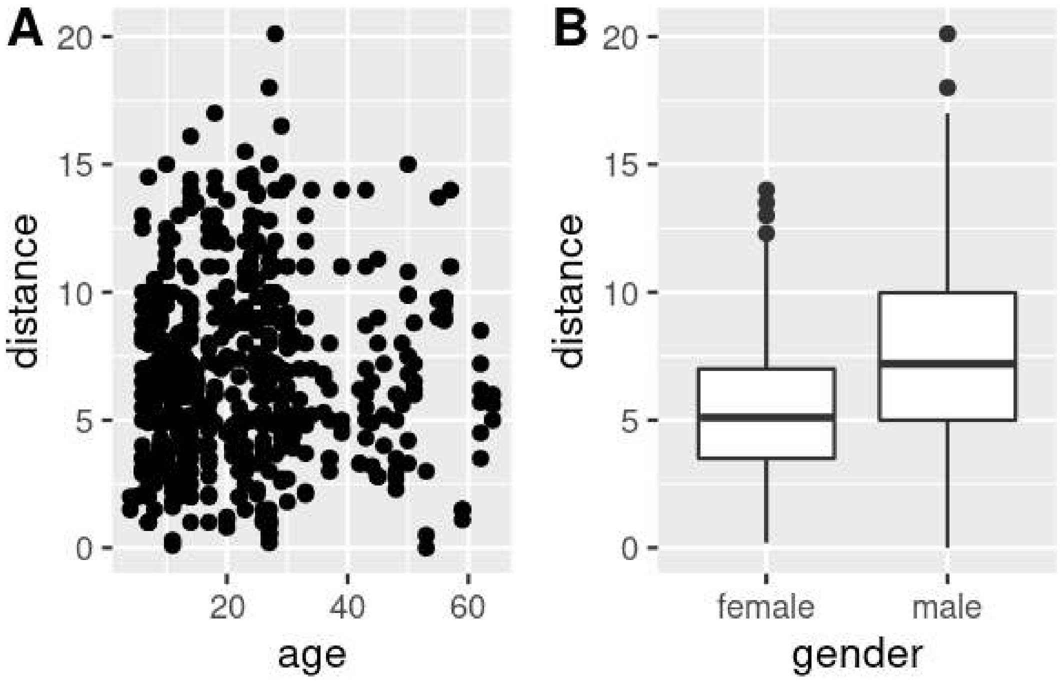
Some publication themes:

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

Composite figures

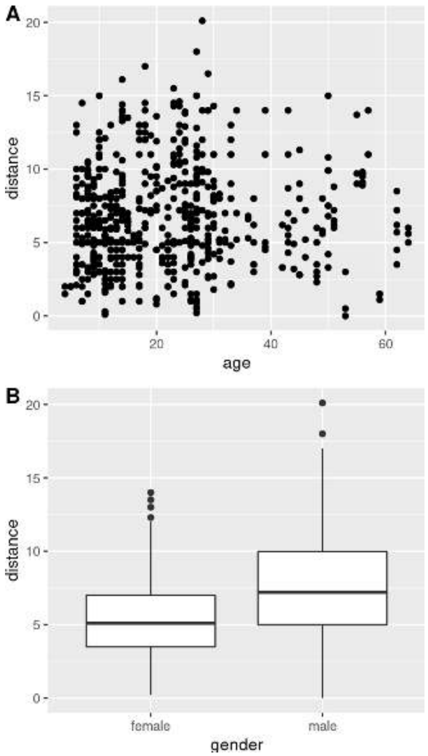
Composite figures: cowplot

```
library(cowplot)
plot1 <- ggplot(paperplanes) + aes(age, distance) + geom_point()
plot2 <- ggplot(paperplanes) + aes(gender, distance) + geom_boxplot()
plot_grid(plot1, plot2, labels = "AUTO")
```

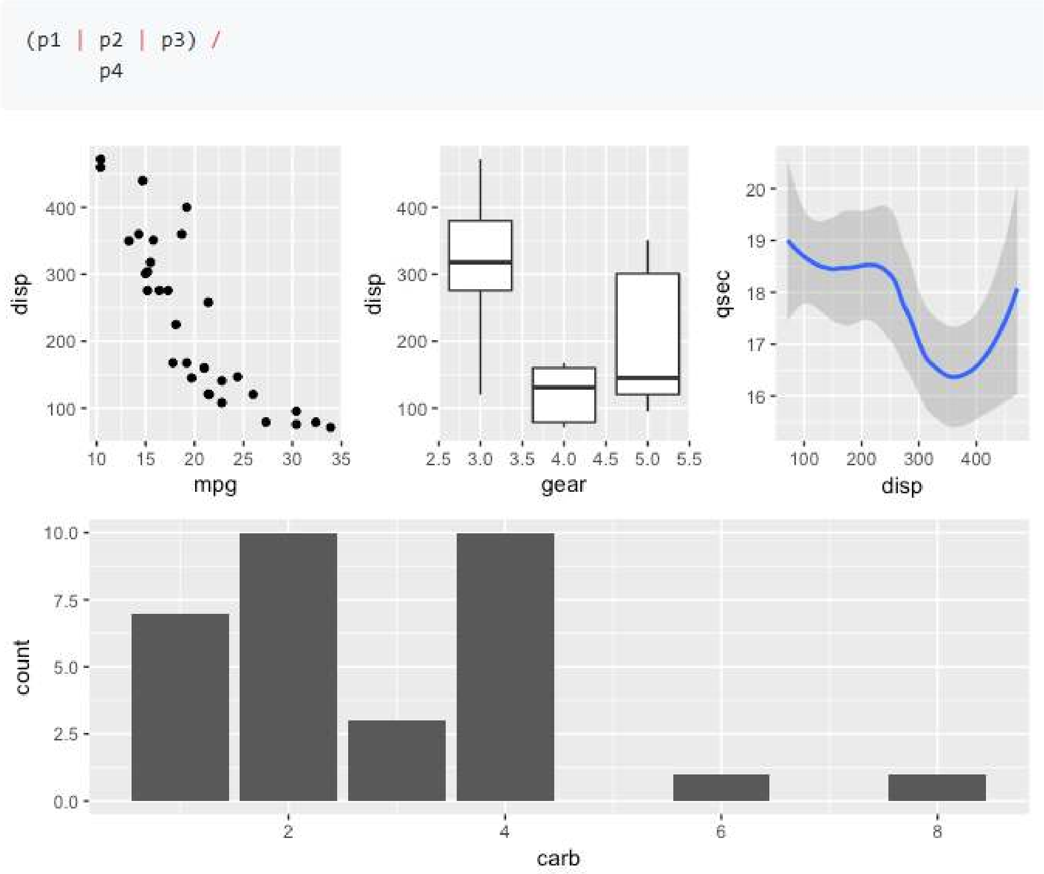


Composite figures

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



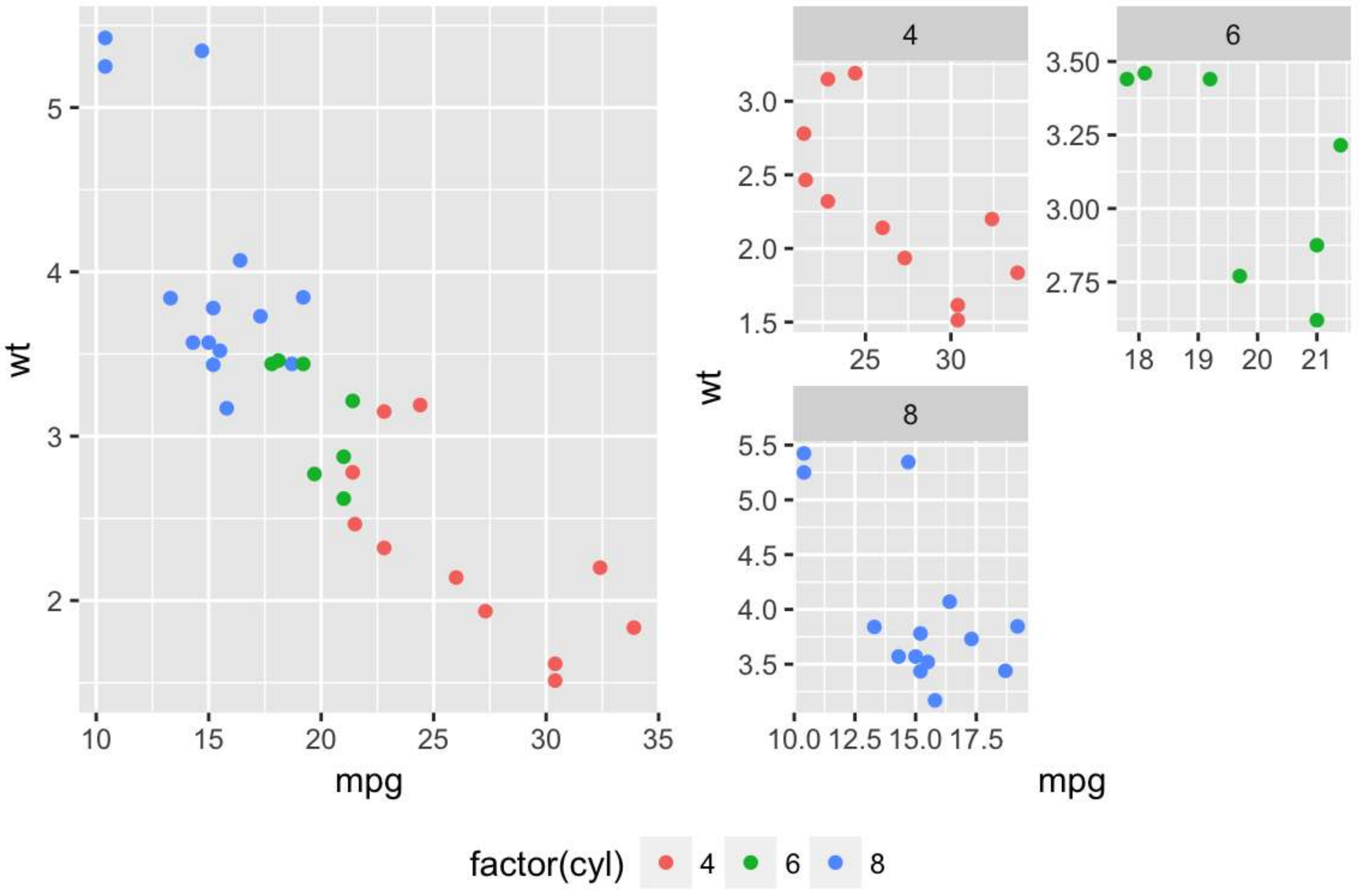
Composite figures: patchwork



```
(p1 | p2 | p3) /
      p4
```

https://github.com/thomasp85/patchwork

Composite figures: egg



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

Saving plot

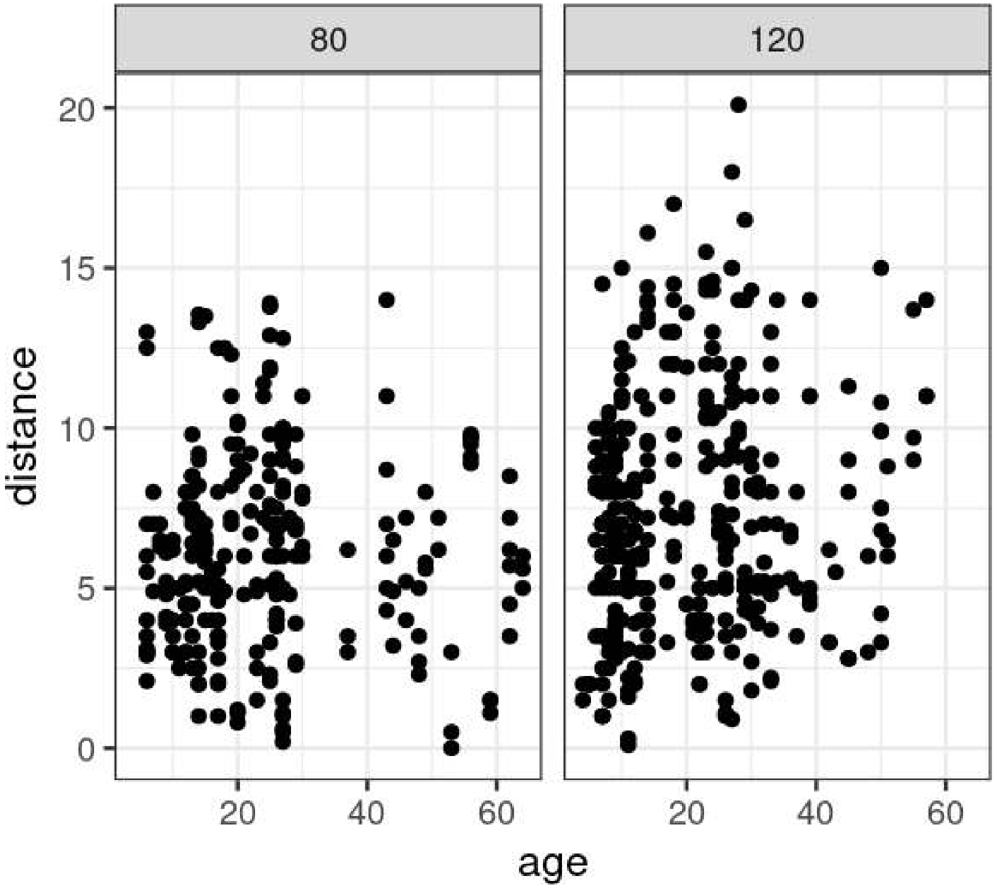
```
ggsave("myplot.pdf")
```

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

Facetting (small multiples)

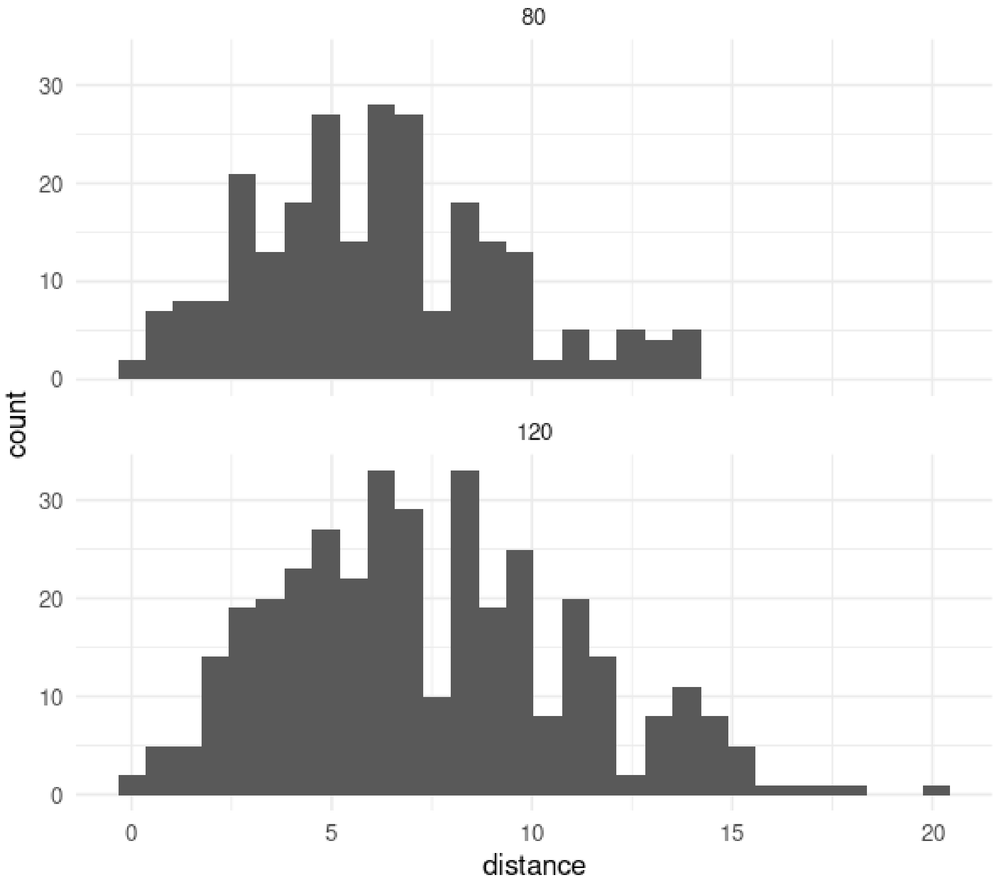
Facetting

```
ggplot(paperplanes) +
  aes(x = age,
      y = distance) +
    geom_point() +
  theme_bw(base_size = 12) +
  facet_wrap(~paper)
```



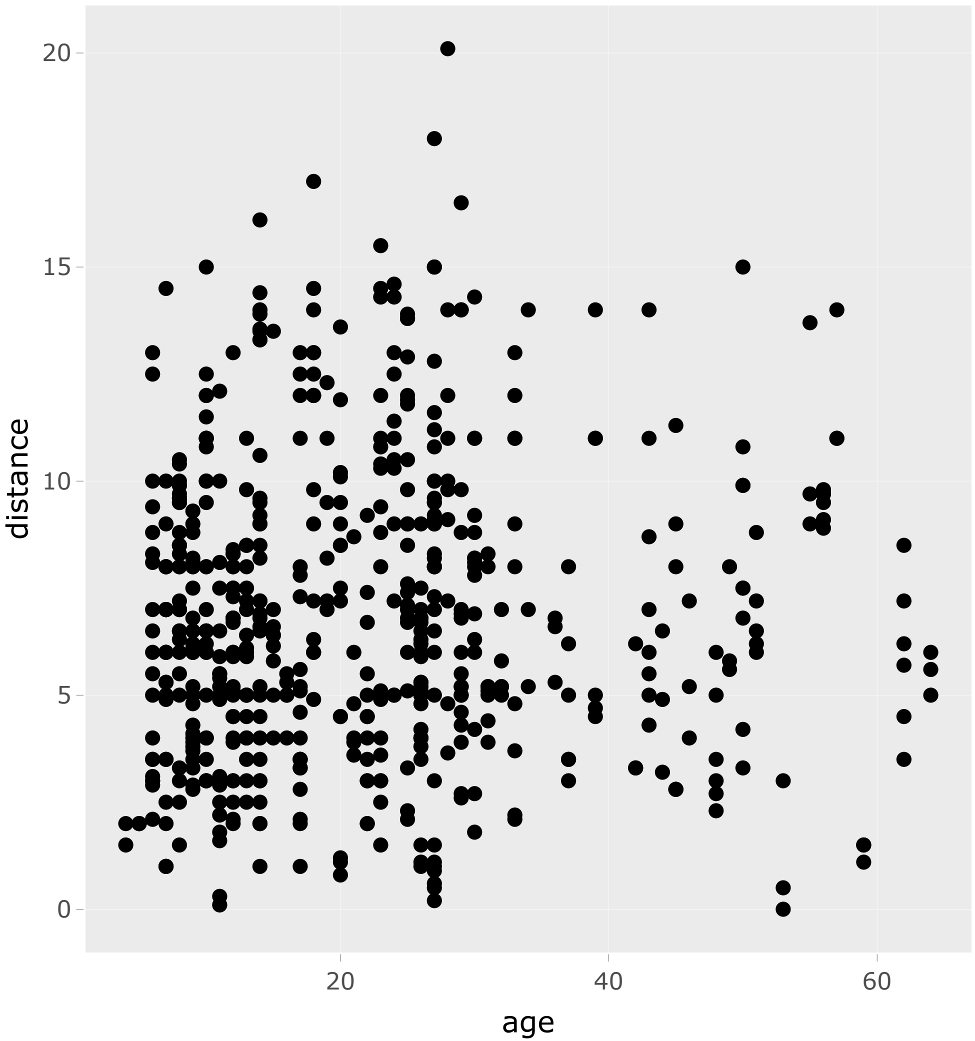
Facetting

```
ggplot(paperplanes) +
  geom_histogram(aes(distance)) +
  theme_minimal(base_size = 8) +
  facet_wrap(~paper, nrow = 2)
```



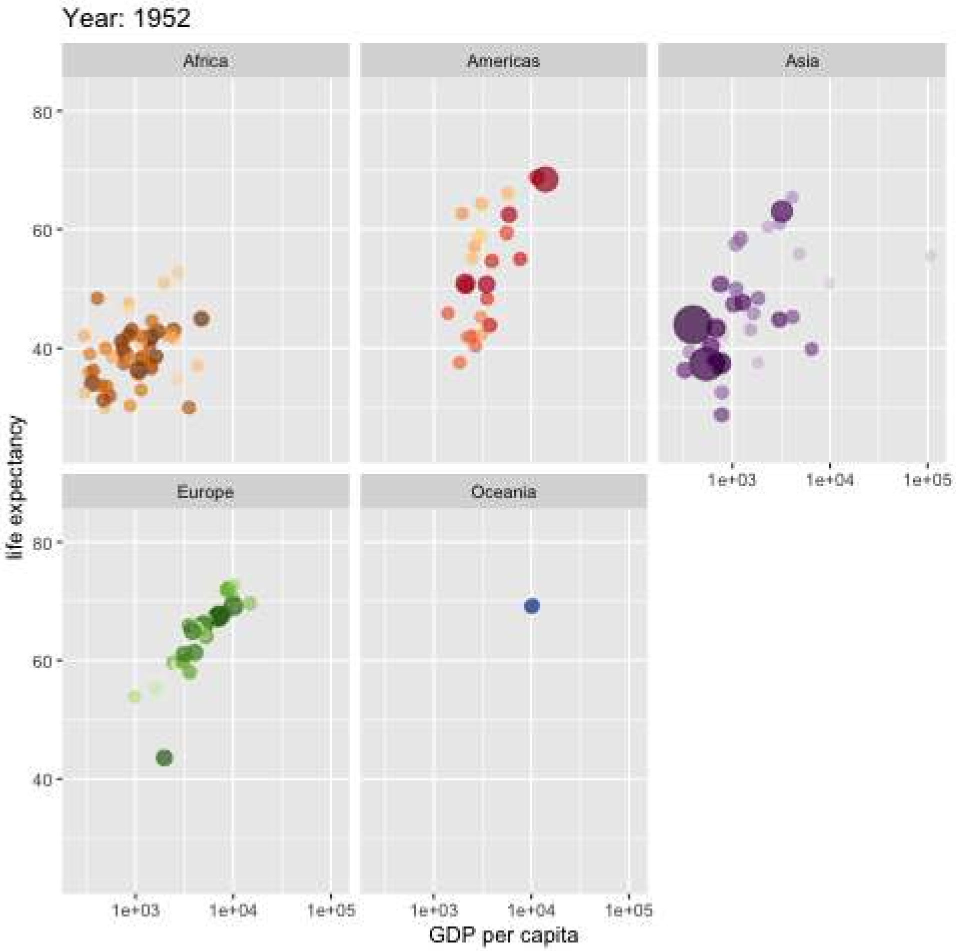
Interactivity: plotly

```r
library(plotly)

myplot <- ggplot(paperplanes) +
  aes(age, distance) +
  geom_point()

ggplotly(myplot)
```

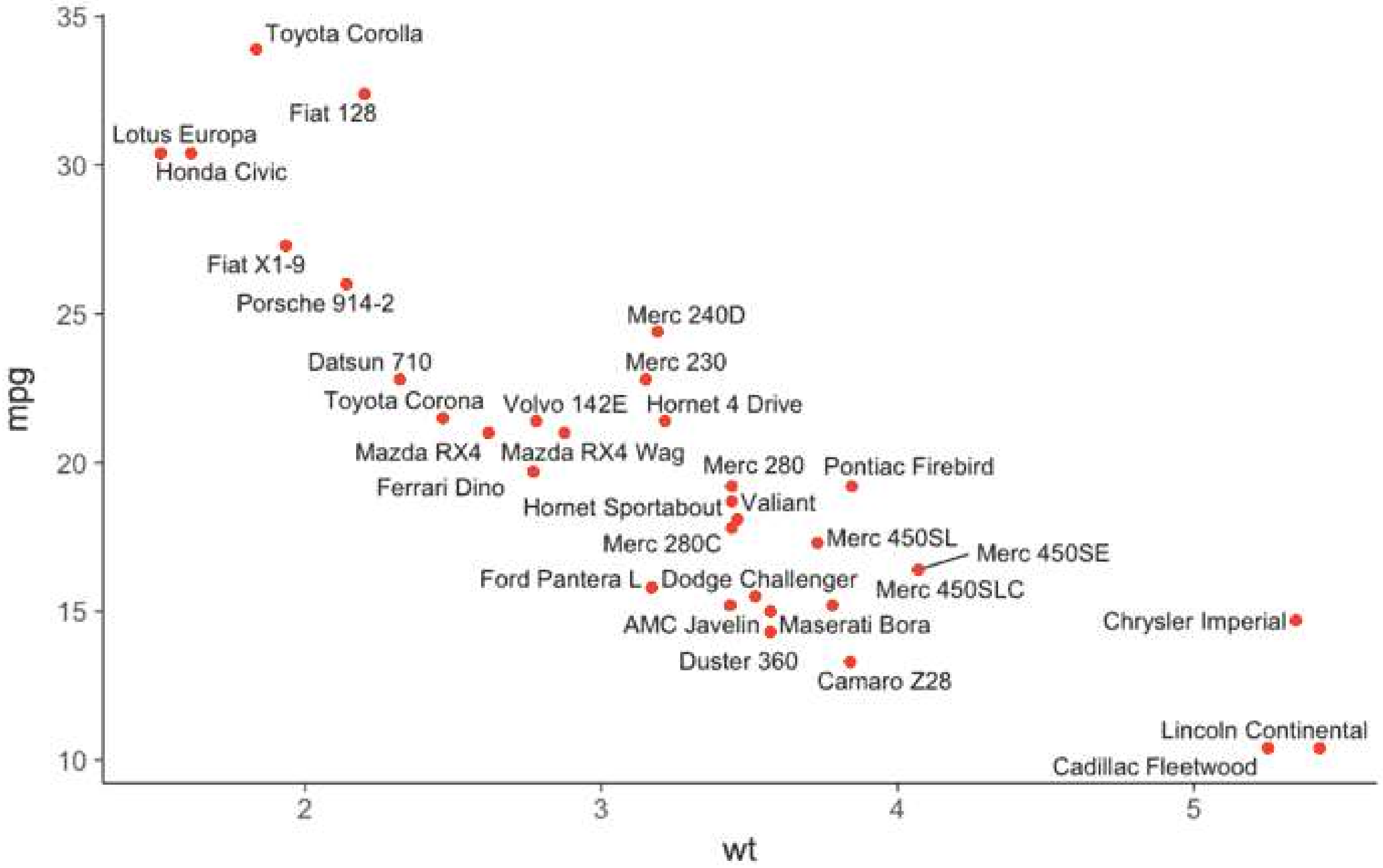


Animated graphs

https://github.com/thomasp85/gganimate



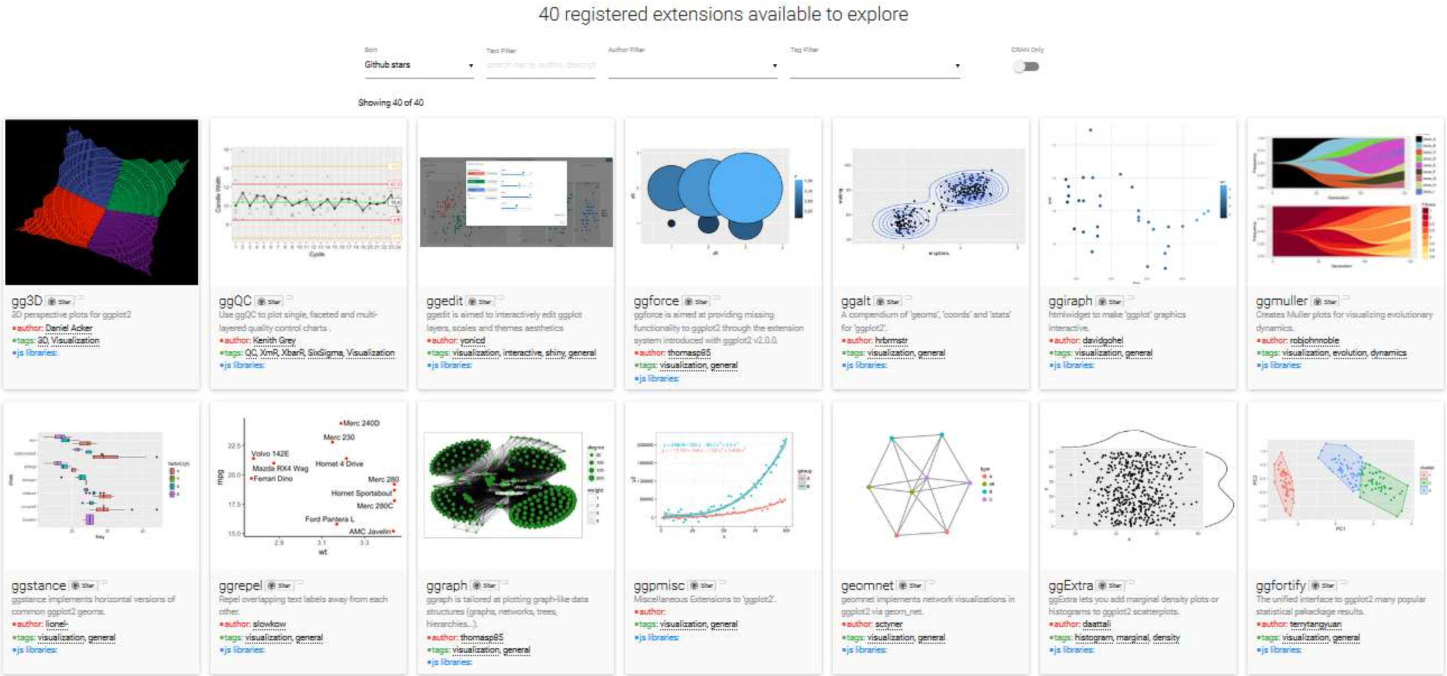
Automatic label placement



https://cran.r-project.org/package=ggrepel

Many extensions!

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



40 registered extensions available to explore

Sort: Github stars

Text Filter: search name, author, descrip

Author Filter

Tag Filter

CRAN Only

Showing 40 of 40

gg3D
3D perspective plots for ggplot2
- author: Daniel Acker
- tags: 3D, Visualization
- js libraries:

ggQC
Use ggQC to plot single, faceted and multi-layered quality control charts .
- author: Kenith Grey
- tags: QC, XmR, XbarR, SixSigma, Visualization
- js libraries:

ggedit
ggedit is aimed to interactively edit ggplot layers, scales and themes aesthetics
- author: yonicd
- tags: visualization, interactive, shiny, general
- js libraries:

ggforce
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
A compendium of 'geoms', 'coords' and 'stats' for 'ggplot2'.
- author: hrbrmstr
- tags: visualization, general
- js libraries:

ggiraph
htmlwidget to make 'ggplot' graphics interactive.
- author: davidgohel
- tags: visualization, general
- js libraries:

ggmuller
Creates Muller plots for visualizing evolutionary dynamics.
- author: robjohnnoble
- tags: visualization, evolution, dynamics
- js libraries:

ggstance
ggstance implements horizontal versions of common ggplot2 geoms.
- author: lionel-
- tags: visualization, general
- js libraries:

ggrepel
Repel overlapping text labels away from each other.
- author: slowkow
- tags: visualization, general
- js libraries:

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

ggpmisc
Miscellaneous Extensions to 'ggplot2'.
- author:
- tags: visualization, general
- js libraries:

geomnet
geomnet implements network visualizations in ggplot2 via geom_net.
- author: sctyner
- tags: visualization, general
- js libraries:

ggExtra
ggExtra lets you add marginal density plots or histograms to ggplot2 scatterplots.
- author: daattali
- tags: histogram, marginal, density
- js libraries:

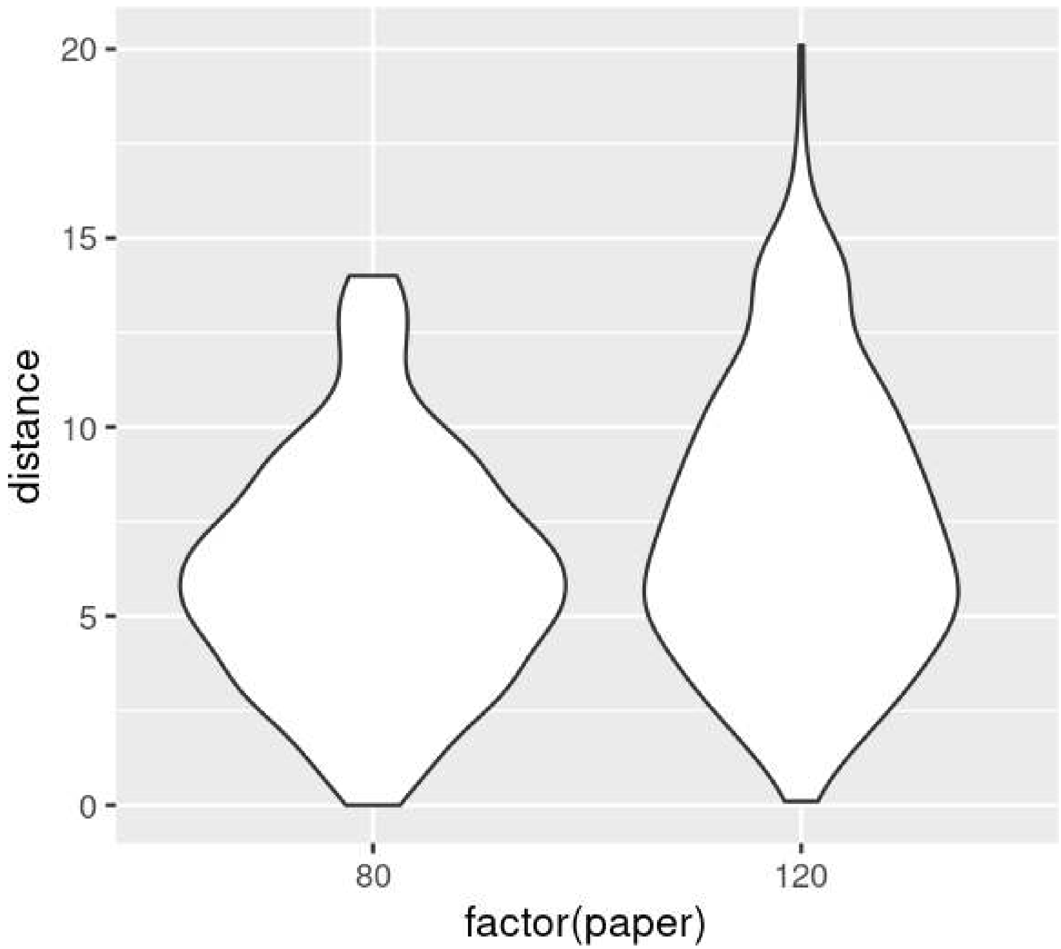
ggfortify
The unified interface to ggplot2 many popular statistical pakackage results.
- author: terrytangyuan
- tags: visualization, general
- js libraries:

Summary

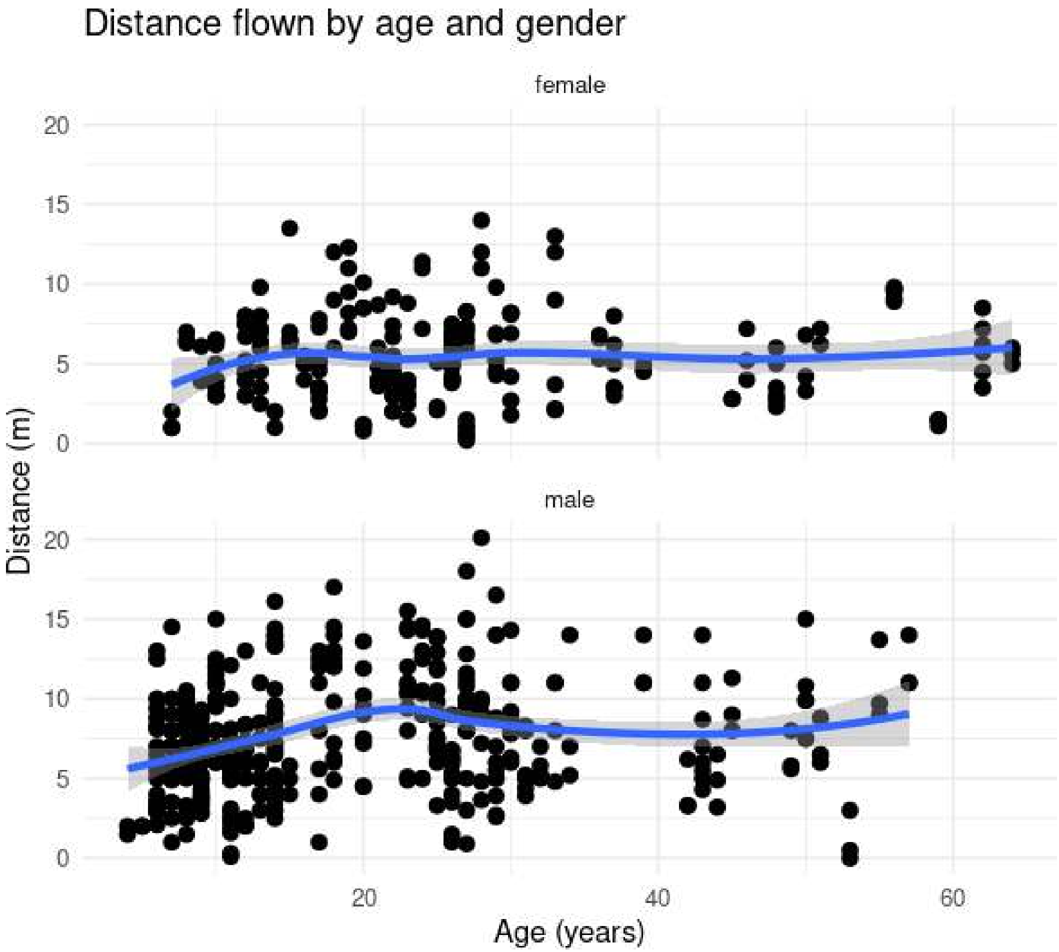
Grammar of graphics

- **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…)

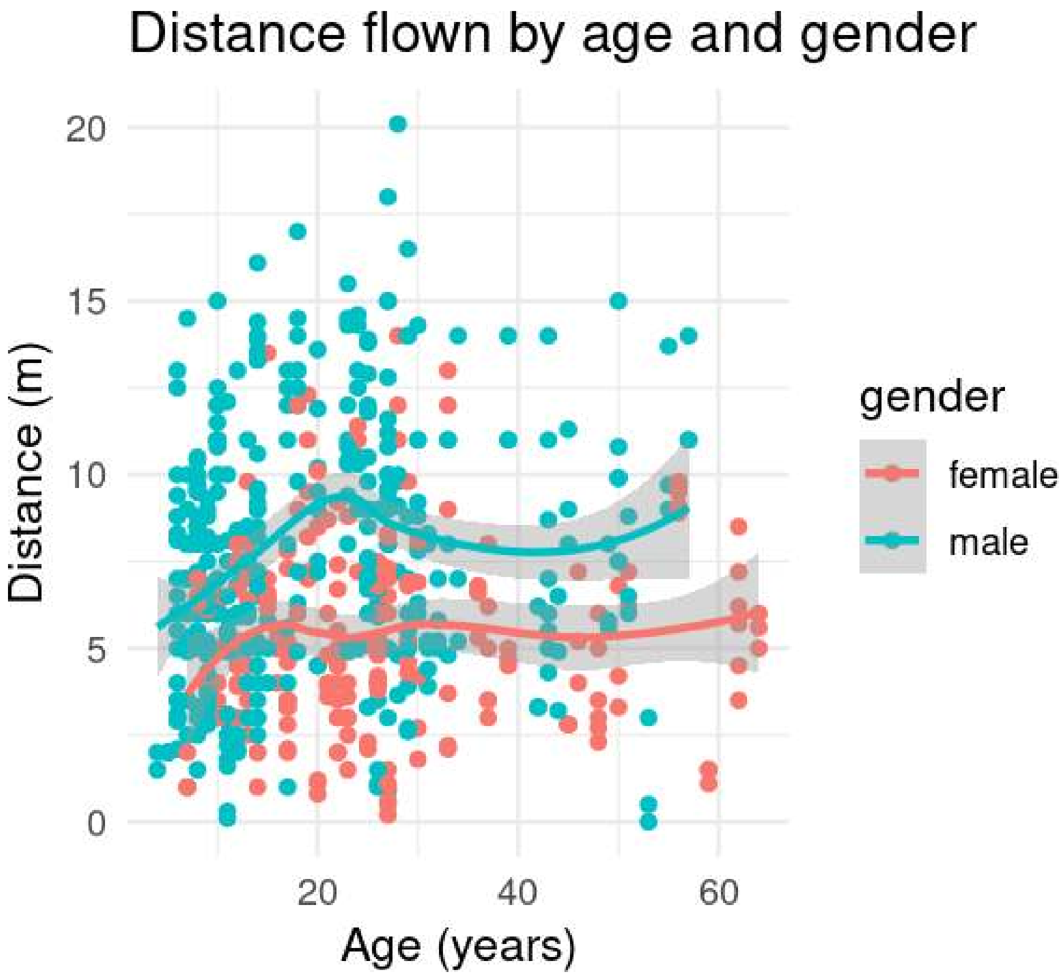
Exercise: make a plot like this one



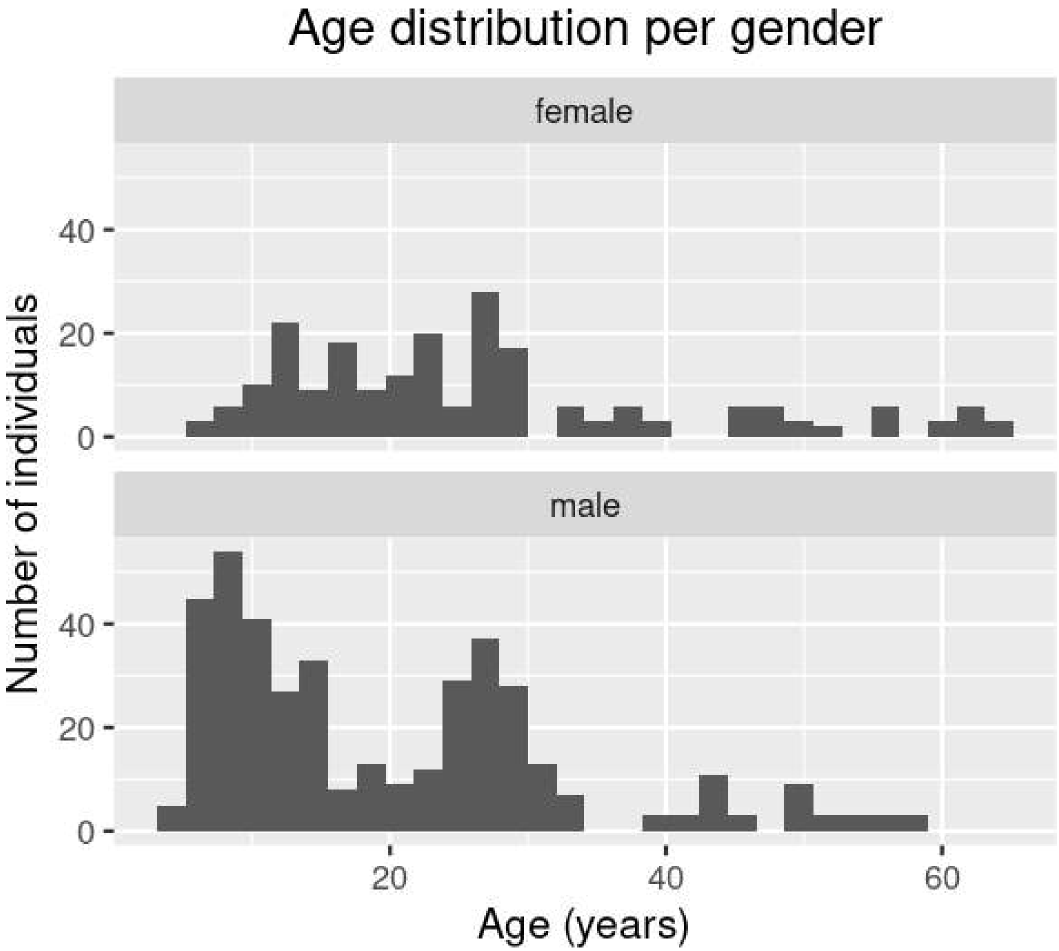
Exercise: make a plot like this one



Exercise: make a plot like this one

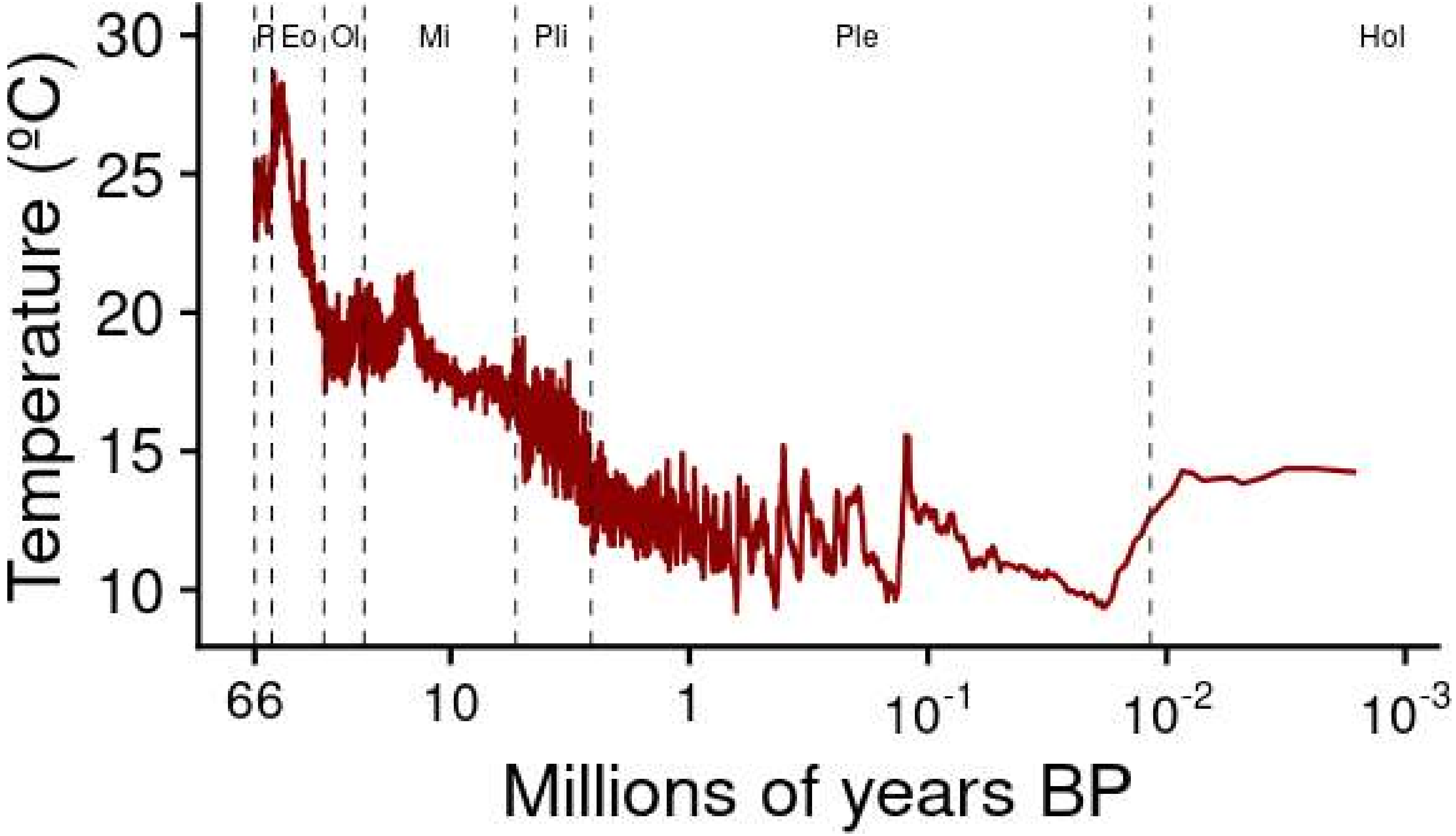


Exercise: make a plot like this one



Exercise: make a plot like this one

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