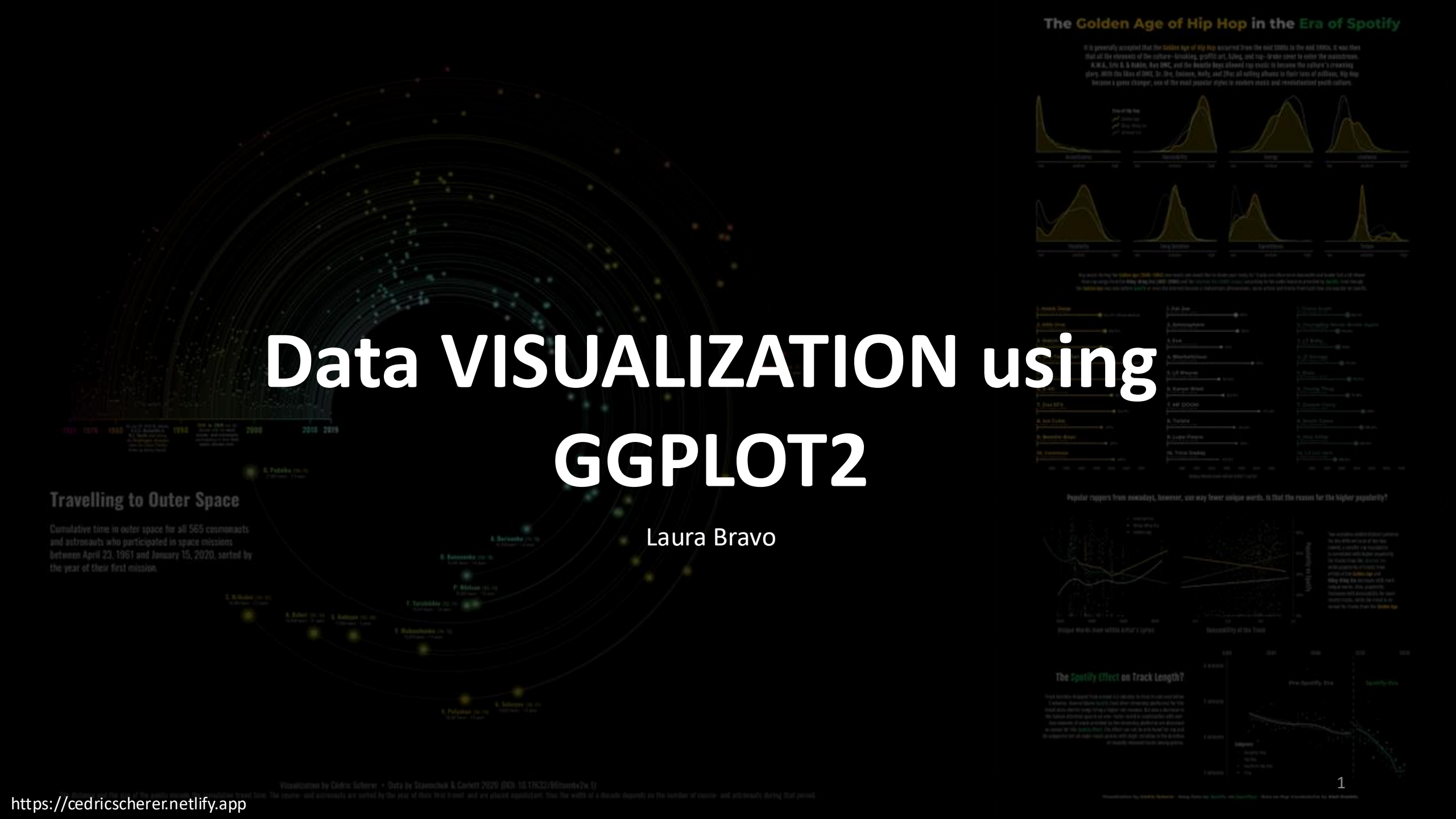


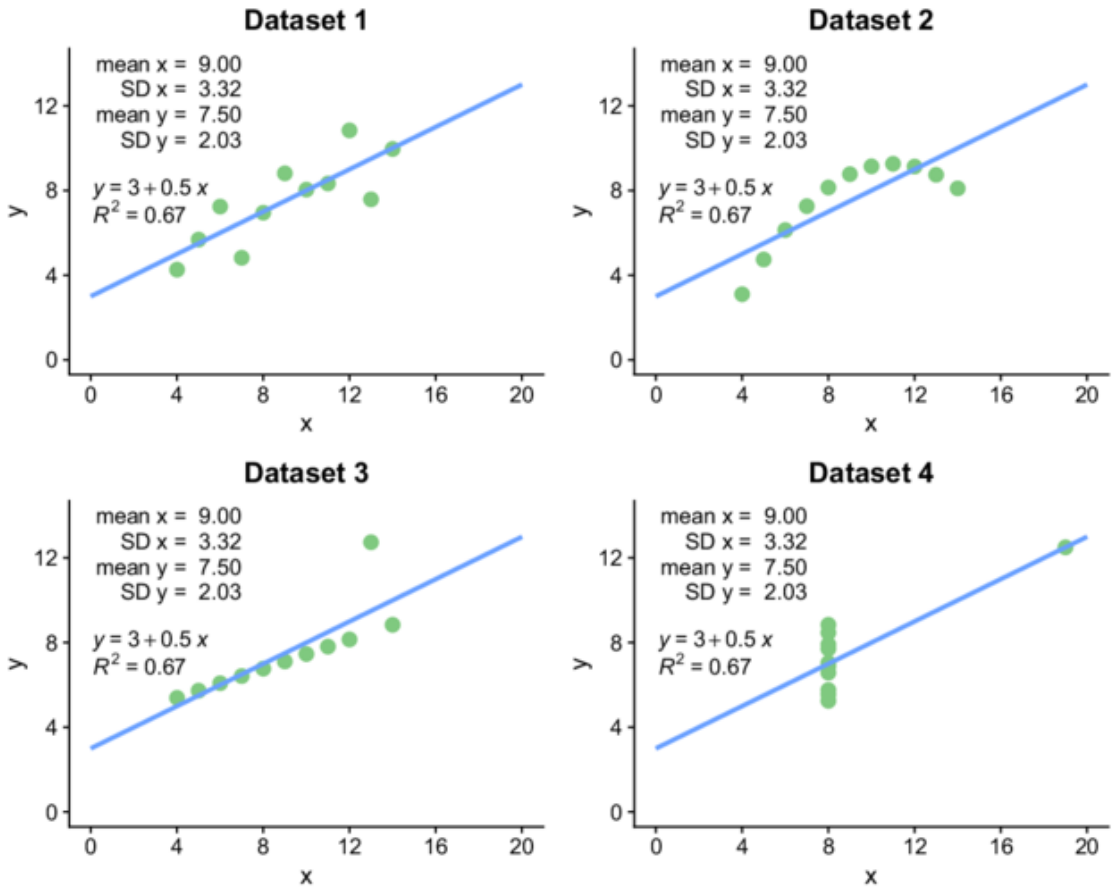
Data VISUALIZATION using GGPLOT2

Laura Bravo

https://cedricscherer.netlify.app

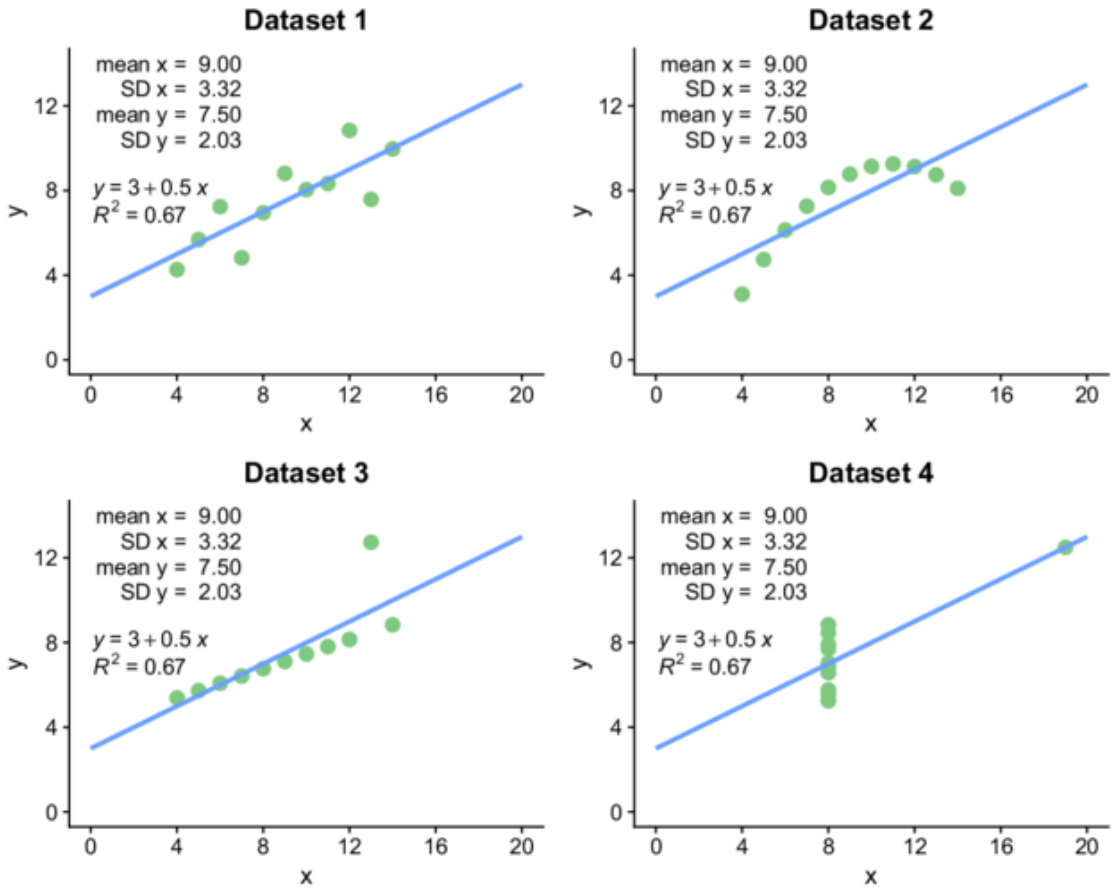
Always, always, always plot the data

Anscombe's quartet (Anscombe, 1973)

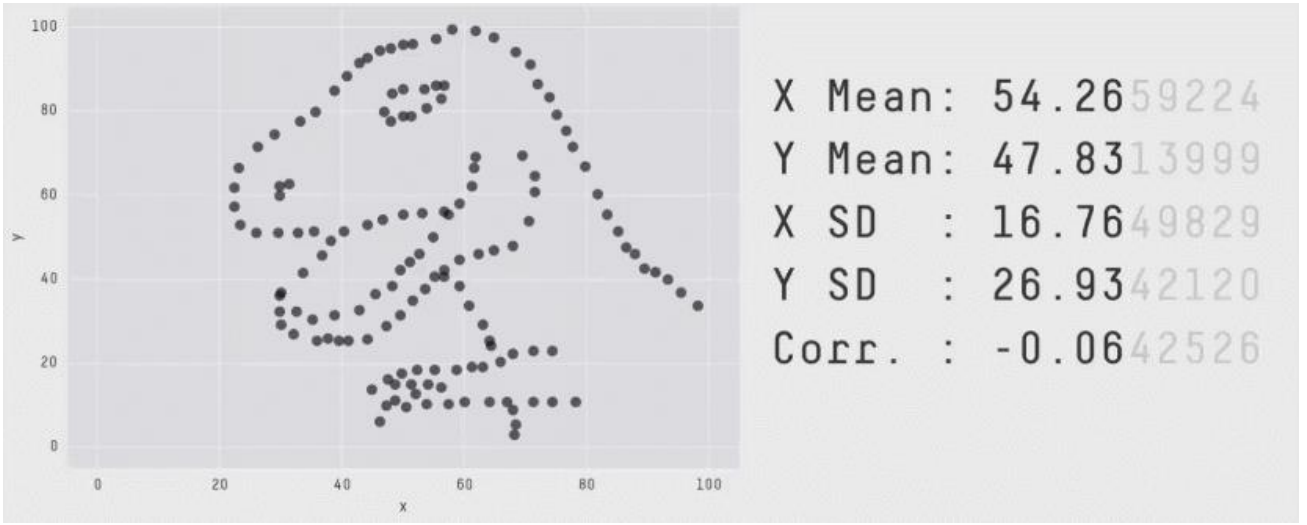


Always, always, always plot the data

Anscombe's quartet (Anscombe, 1973)



Datasaurus dozen (Matejka & Fitzmaurice, 2017)



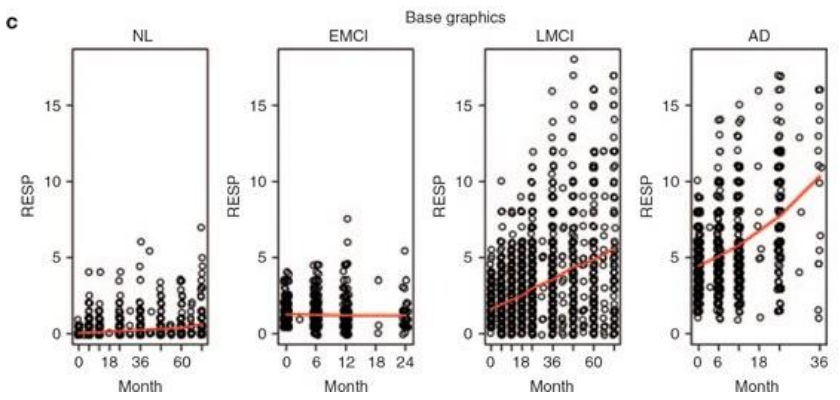
A picture is worth more than 1000 words! (Anonymous)

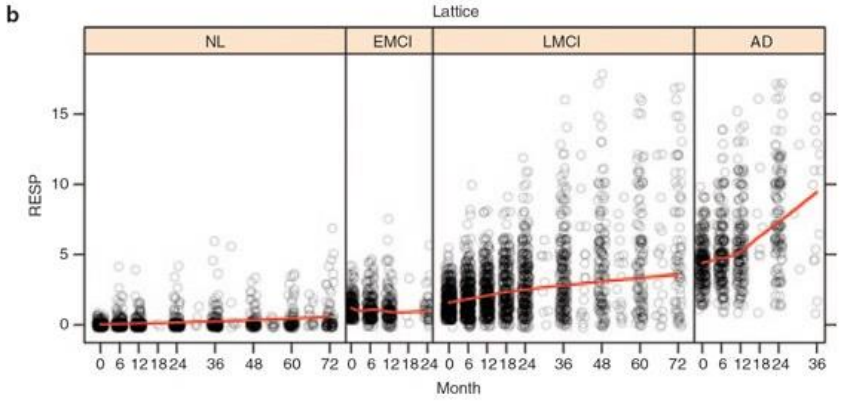
Graphics in R

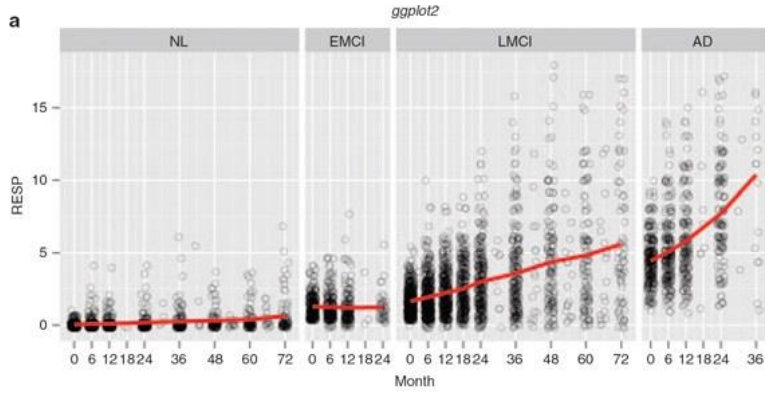
Base R (1993)

Lattice (1993) (2008)
Trellis graphics – Bell Laboratories

Ggplot2 (2005)
Layered grammar of graphics







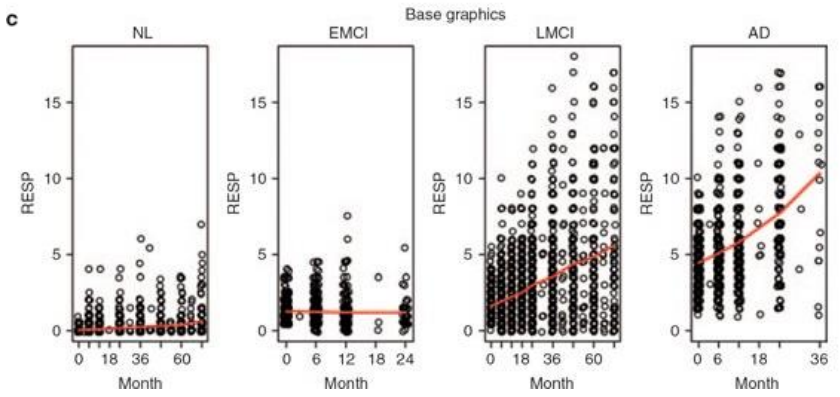
Application of ggplot2 to Pharmacometric Graphics (Kaori Ito, 2013)

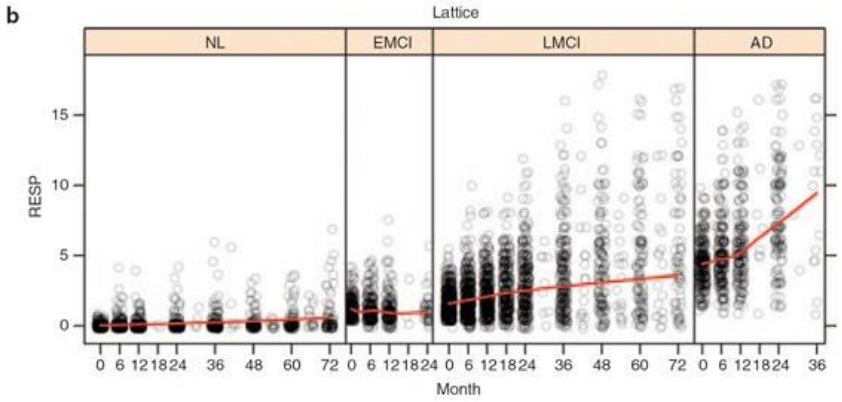
Graphics in R

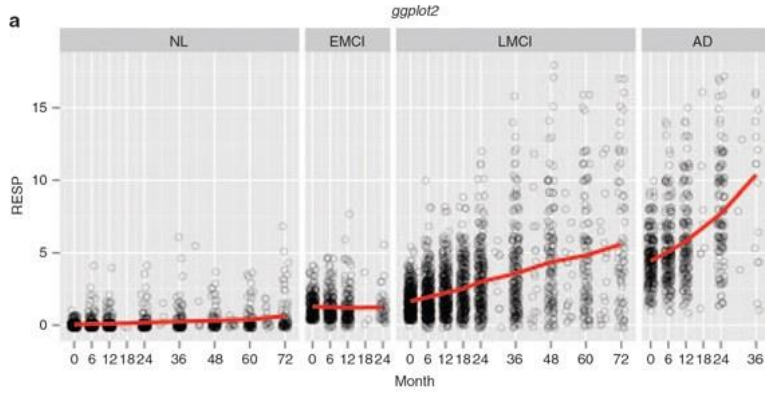
Base R (1993)

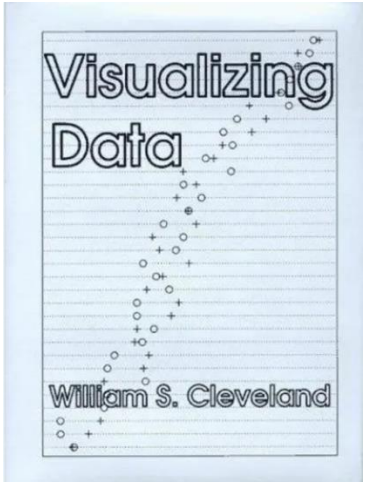
Lattice (1993) (2008)
Trellis graphics – Bell Laboratories

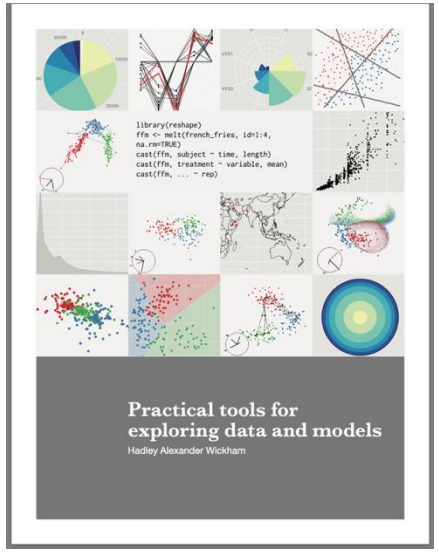
Ggplot2 (2005)
Layered grammar of graphics











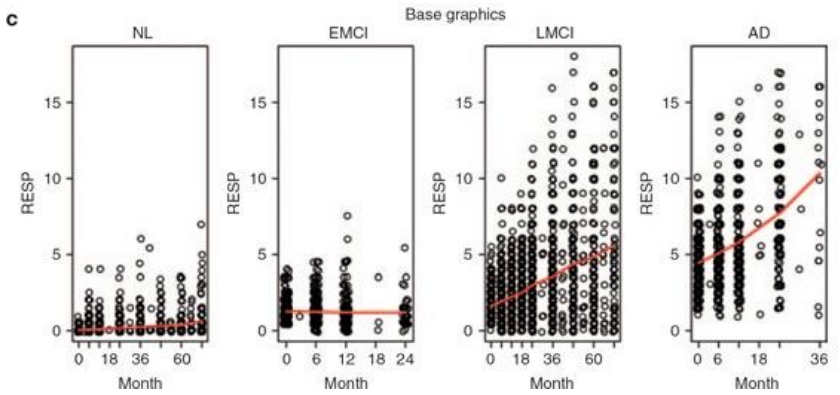
Application of ggplot2 to Pharmacometric Graphics (Kaori Ito, 2013)

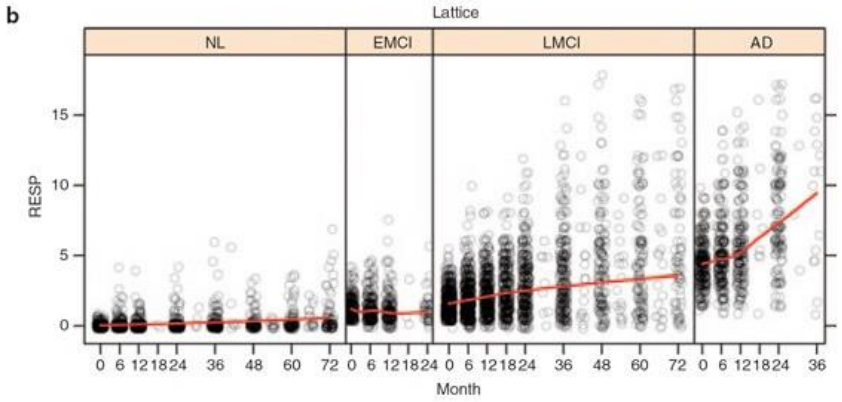
Graphics in R

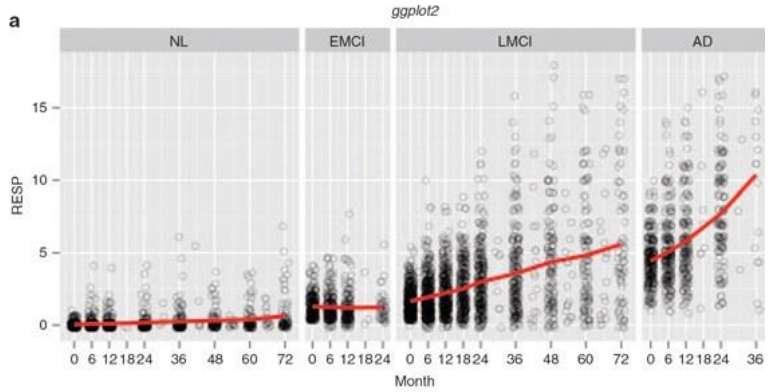
Base R (1993)

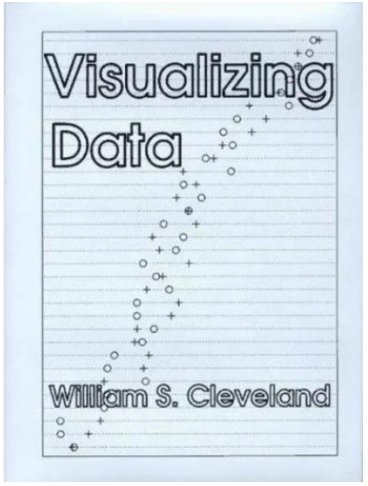
Lattice (1993) (2008)
Trellis graphics – Bell Laboratories

Ggplot2 (2005)
Layered grammar of graphics

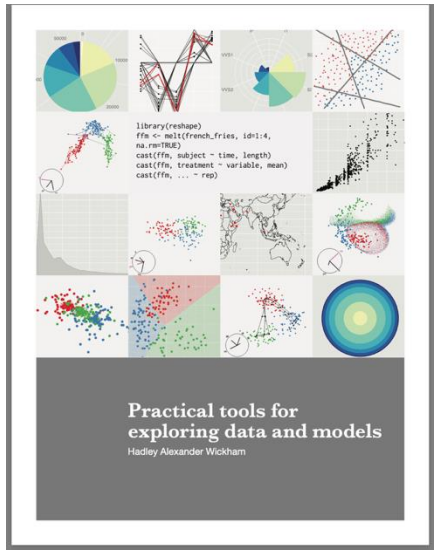








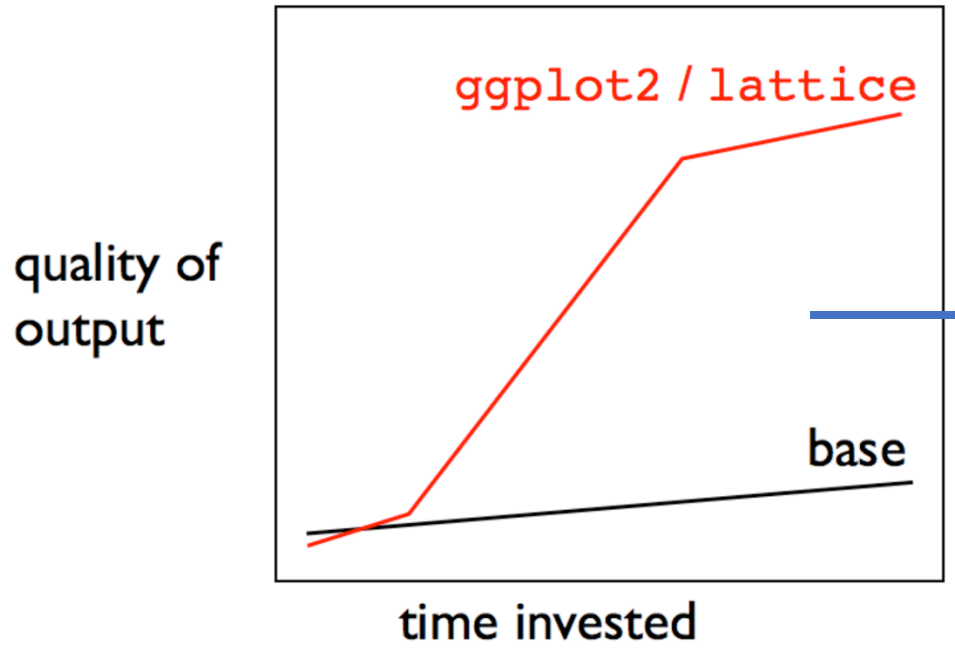
Hadley Wickham



Application of ggplot2 to Pharmacometric Graphics (Kaori Ito, 2013)

Why use ggplot2?

week one ....



* figure is totally fabricated but, I claim, s

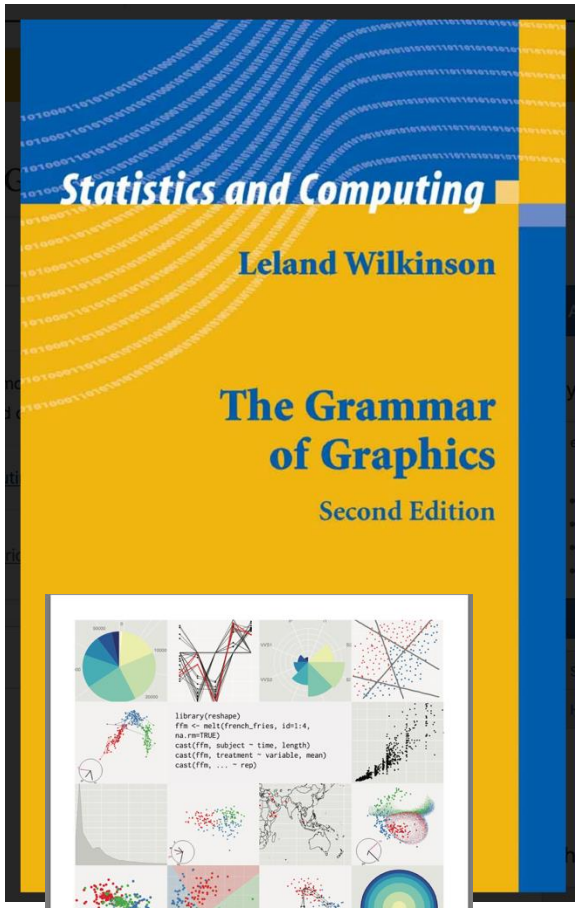
after you've climbed the steepest part of the learning curve ...

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

How does it work? Grammar of graphics

Describe the fundamental features that underlie all statistical graphics.
Answer question - what is a statistical graphic?

2005



Idea: You can build every graph from same components

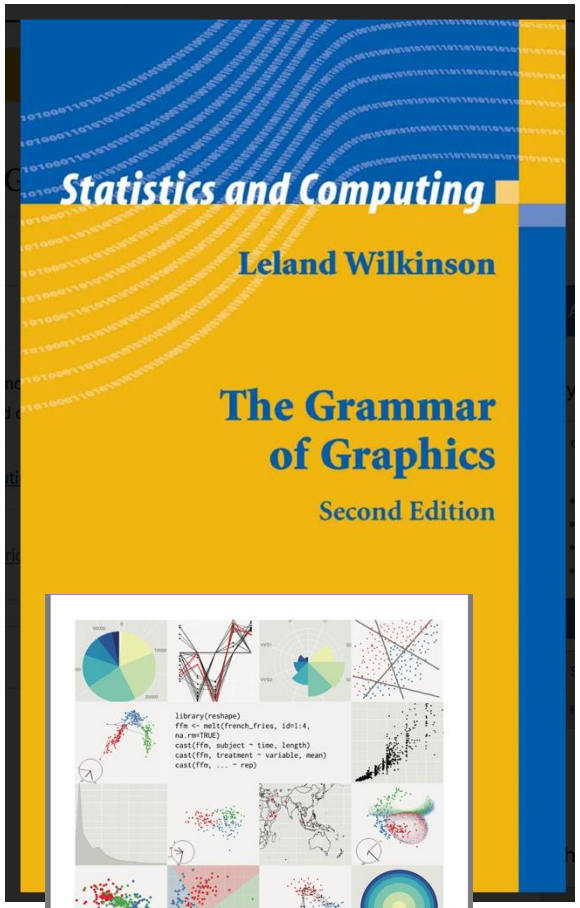
Data: The raw data that you want to plot.
Geometries geom_: The geometric shapes that will represent the data.
Aesthetics aes(): What variables we want plotted and aesthetics of the geometric and statistical objects, such as position, color, size, shape, and transparency

Hadley Wickham

How does it work? Grammar of graphics

Describe the fundamental features that underlie all statistical graphics.
Answer question - what is a statistical graphic?

2005



Idea: You can build every graph from same components

Data: The raw data that you want to plot.
Geometries geom_: The geometric shapes that will represent the data.
Aesthetics aes(): What variables we want plotted and aesthetics of the geometric and statistical objects, such as position, color, size, shape, and transparency

Scales scale_.
Statistical transformations stat_:
Coordinate system coord_:
Facets facet_:
Visual themes theme():

Hadley Wickham

How does it work? Grammar of graphics

Idea: You can build every graph from same components

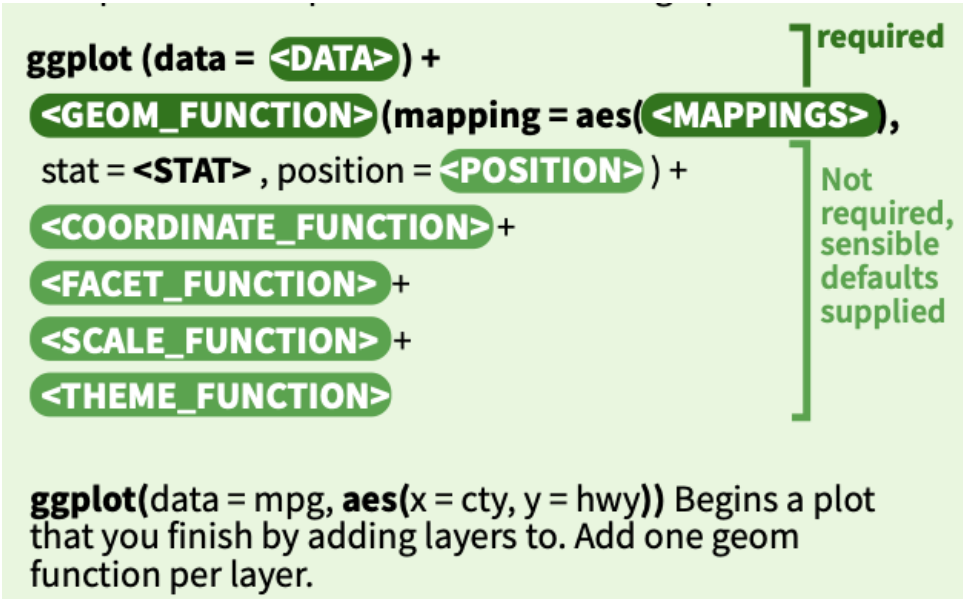
Data: The raw data that you want to plot.
Geometries geom_: The geometric shapes that will represent the data.
Aesthetics aes(): Aesthetics of the geometric and statistical objects, such as position, color, size, shape, and transparency

+

Scales scale_.
Statistical transformations stat_:
Coordinate system coord_:
Facets facet_:
Visual themes theme():

Layers!!!

In Practice:



How does it work?

```
library(ggplot2)

mpg
## # A tibble: 234 × 11
##    manufacturer      model displ  year   cyl      trans   drv   cty
##           <chr>      <chr> <dbl> <int> <int>      <chr> <chr> <int> <i
## 1          audi         a4   1.8  1999     4   auto(l5)     f    18
## 2          audi         a4   1.8  1999     4 manual(m5)     f    21
## 3          audi         a4   2.0  2008     4 manual(m6)     f    20
## 4          audi         a4   2.0  2008     4   auto(av)     f    21
## 5          audi         a4   2.8  1999     6   auto(l5)     f    16
## 6          audi         a4   2.8  1999     6 manual(m5)     f    18
## 7          audi         a4   3.1  2008     6   auto(av)     f    18
## 8          audi a4 quattro   1.8  1999     4 manual(m5)     4    18
## 9          audi a4 quattro   1.8  1999     4   auto(l5)     4    16
## 10         audi a4 quattro   2.0  2008     4 manual(m6)     4    20
## # ... with 224 more rows, and 2 more variables: fl <chr>, class <chr>
```

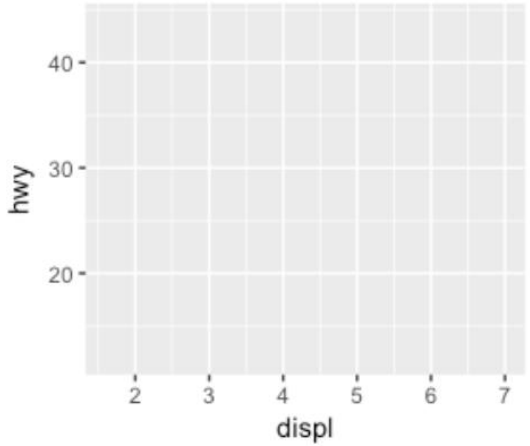
https://uc-r.github.io/ggplot_intro

How does it work?

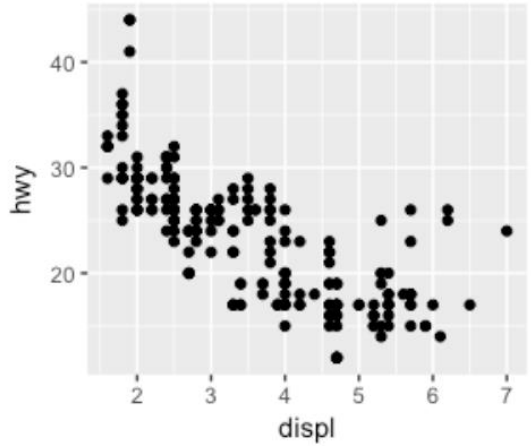
```
# create canvas
ggplot(mpg)

# variables of interest mapped
ggplot(mpg, aes(x = displ, y = hwy))

# data plotted
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point()
```

Canvas

Canvas + variables mapped to axes



Data plotted



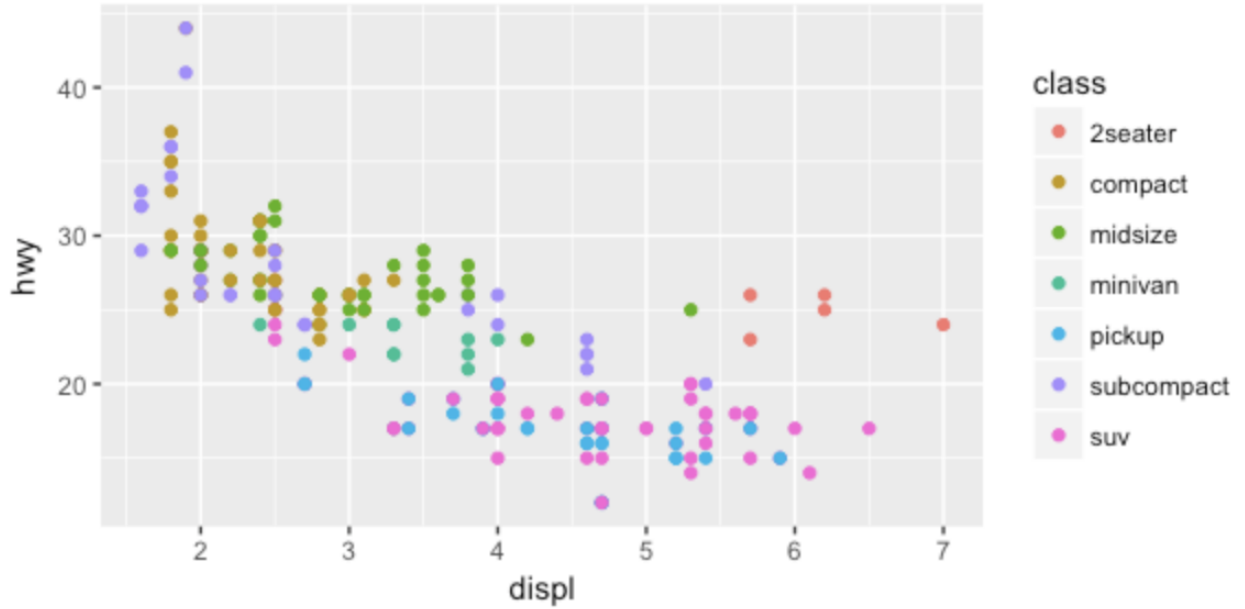
Layers!!!

https://uc-r.github.io/ggplot_intro

How does it work? Aesthetic mappings

Properties of the data and use them to influence visual characteristics, such as *position*, *color*, *size*, *shape*, or *transparency*. Each visual characteristic can thus encode an aspect of the data and be used to convey information.

```
ggplot(mpg, aes(x = displ, y = hwy, color = class)) +
  geom_point()
```



https://uc-r.github.io/ggplot_intro

How does it work? Geometric shapes

Each of these geometries will leverage the aesthetic mappings supplied although the specific visual properties that the data will map to will vary.

Geometries geom_: The geometric shapes that will represent the data.

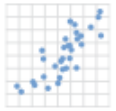
Fundamental building blocks

TWO VARIABLES

continuous x , continuous y

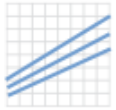
e <- ggplot(mpg, aes(cty, hwy))

e + geom_label(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)** x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust



e + geom_jitter(height = 2, width = 2**)** x, y, alpha, color, fill, shape, size

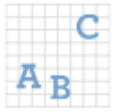
e + geom_point(), x, y, alpha, color, fill, shape, size, stroke



e + geom_quantile(), x, y, alpha, color, group, linetype, size, weight

e + geom_rug(sides = "bl"**)**, x, y, alpha, color, linetype, size

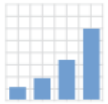
e + geom_smooth(method = lm**)**, x, y, alpha, color, fill, group, linetype, size, weight



e + geom_text(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)**, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

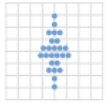
discrete x , continuous y

f <- ggplot(mpg, aes(class, hwy))



f + geom_col(), x, y, alpha, color, fill, group, linetype, size

f + geom_boxplot(), x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight



f + geom_dotplot(binaxis = "y", stackdir = "center"**)**, x, y, alpha, color, fill, group

f + geom_violin(scale = "area"**)**, x, y, alpha, color, fill, group, linetype, size, weight

Ggplot2 cheatsheet

How does it work? Geometric shapes

Geometries geom_: The geometric shapes that will represent the data.

Fundamental building blocks

scatterplot

barchart

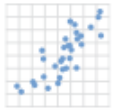
Box-and-whisker plot

TWO VARIABLES

continuous x , continuous y

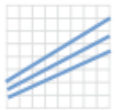
e <- ggplot(mpg, aes(cty, hwy))

e + geom_label(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)** x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust



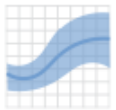
e + geom_jitter(height = 2, width = 2**)** x, y, alpha, color, fill, shape, size

e + geom_point(), x, y, alpha, color, fill, shape, size, stroke



e + geom_quantile(), x, y, alpha, color, group, linetype, size, weight

e + geom_rug(sides = "bl"**)**, x, y, alpha, color, linetype, size

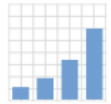


e + geom_smooth(method = lm**)**, x, y, alpha, color, fill, group, linetype, size, weight

e + geom_text(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE**)**, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

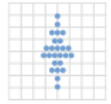
discrete x , continuous y

f <- ggplot(mpg, aes(class, hwy))



f + geom_col(), x, y, alpha, color, fill, group, linetype, size

f + geom_boxplot(), x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight



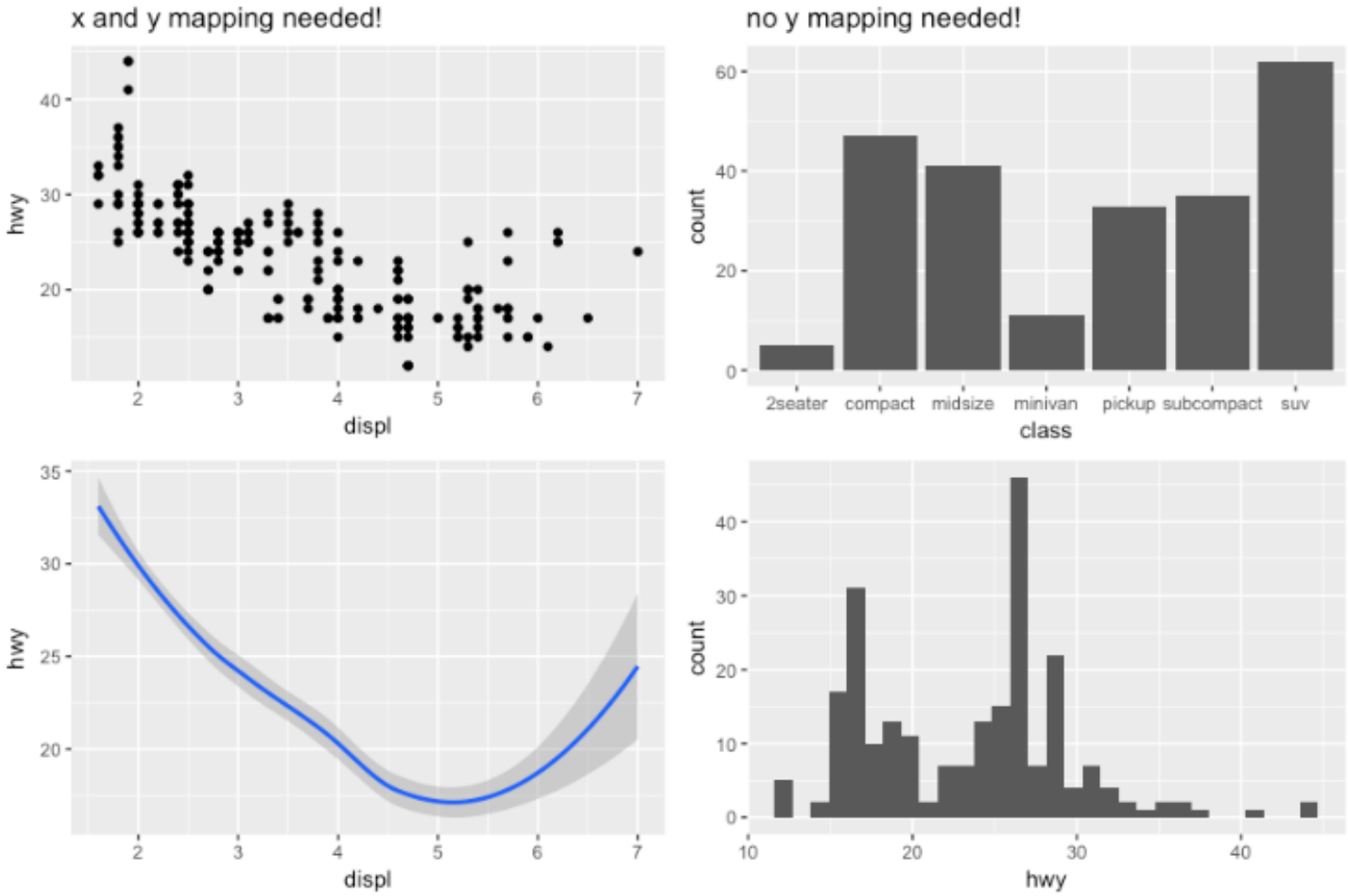
f + geom_dotplot(binaxis = "y", stackdir = "center"**)**, x, y, alpha, color, fill, group

f + geom_violin(scale = "area"**)**, x, y, alpha, color, fill, group, linetype, size, weight

Ggplot2 cheatsheet

How does it work? Geometric shapes

```
# Left column: x and y mapping needed!
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point()

ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_smooth()

# Right column: no y mapping needed!
ggplot(data = mpg, aes(x = class)) +
  geom_bar()

ggplot(data = mpg, aes(x = hwy)) +
  geom_histogram()
```



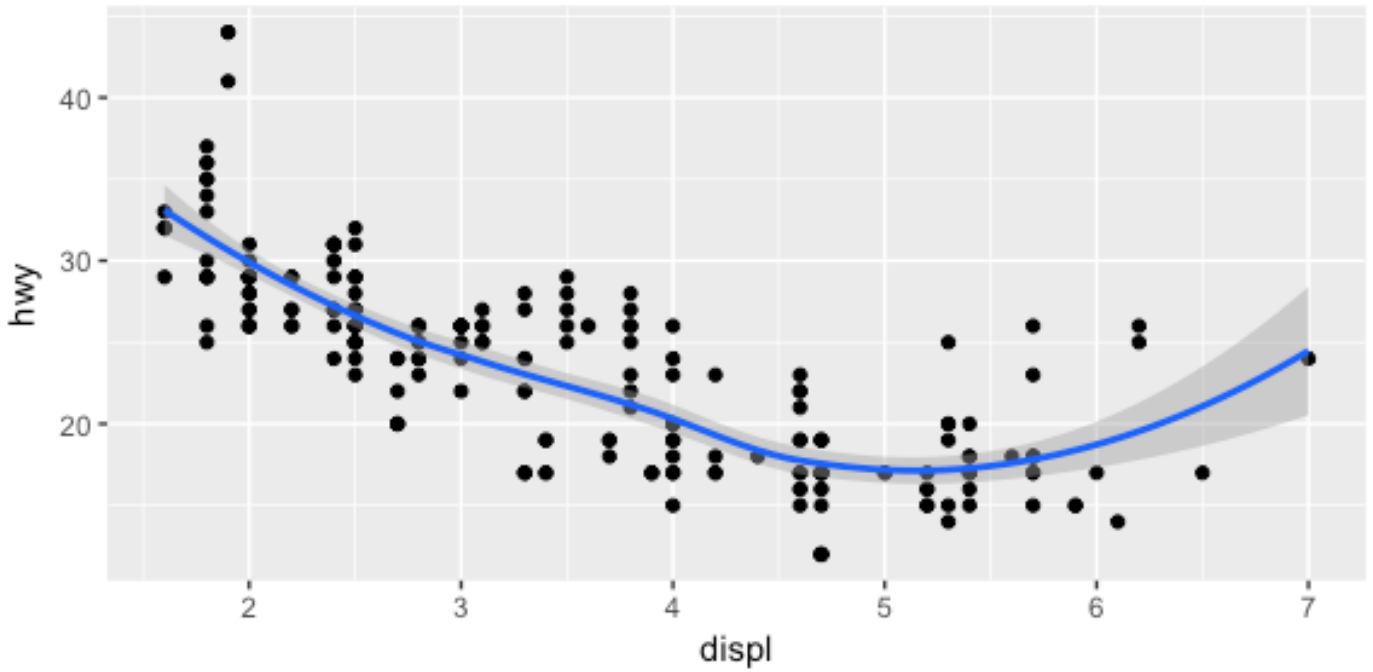
https://uc-r.github.io/ggplot_intro

How does it work? Geometric shapes

What makes this really powerful is that you can add ***multiple*** geometries to a plot, thus allowing you to create complex graphics showing multiple aspects of your data.

Each function returns a layer!!

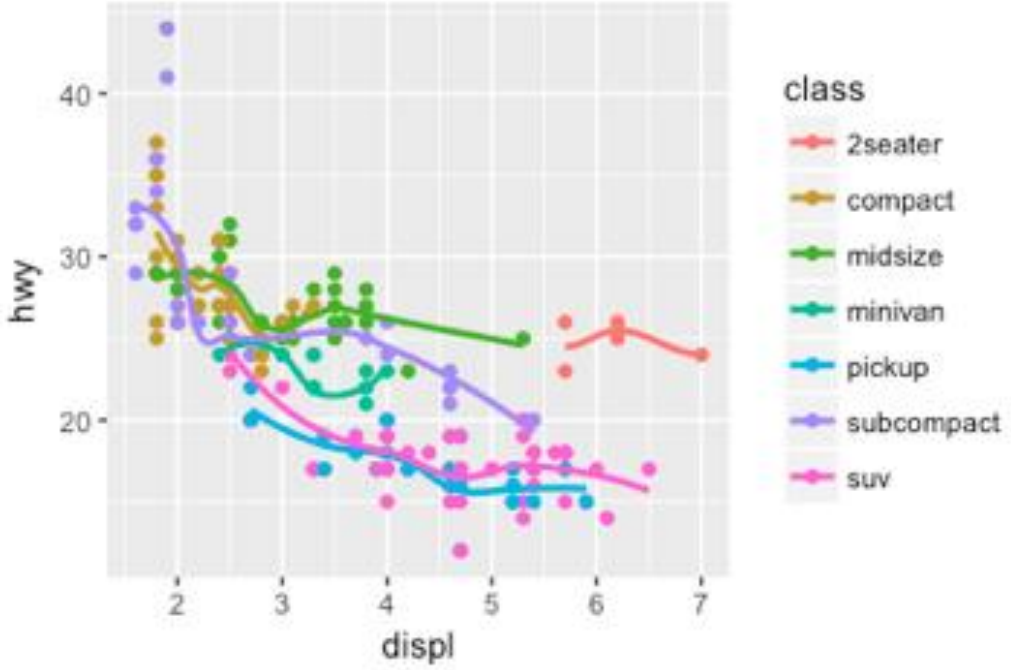
```
# plot with both points and smoothed line
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth()
```

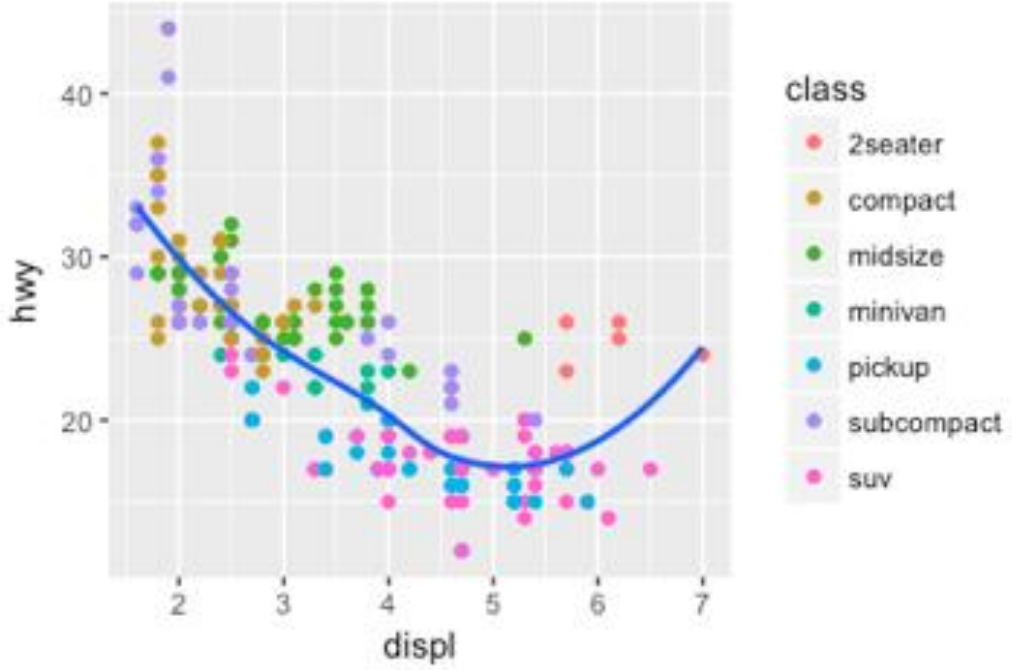


https://uc-r.github.io/ggplot_intro

How does it work? Geometric shapes and aesthetic mappings

If we specify an aesthetic within ggplot it will be passed on to each geom that follows. Or we can specify certain aes within each geom, which allows us to only show certain characteristics for that specificy layer (i.e. geom_point).





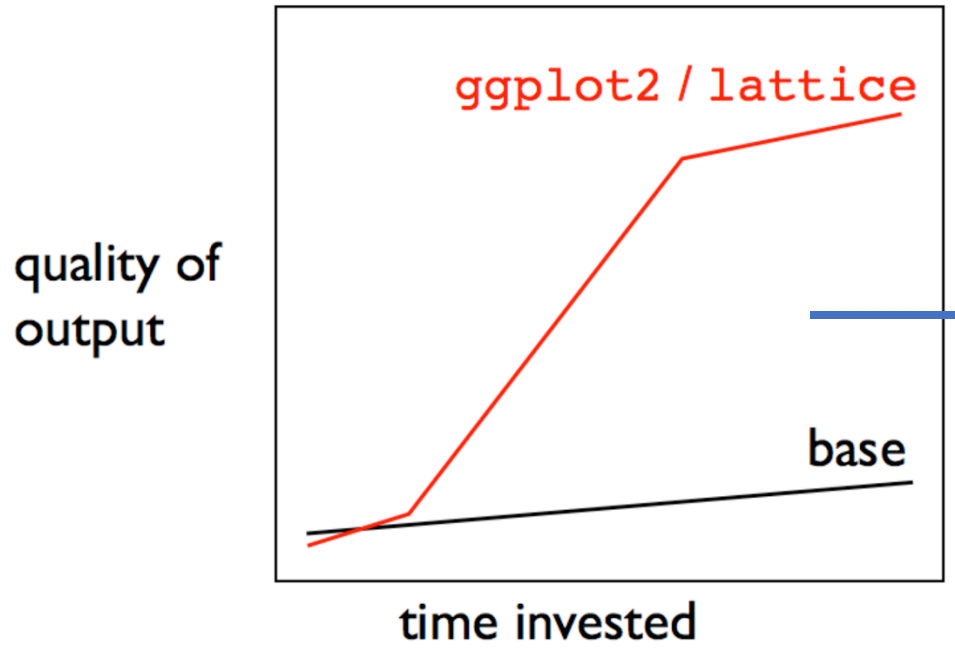

```
# color aesthetic passed to each geom layer
ggplot(mpg, aes(x = displ, y = hwy, color = class)) +
  geom_point() +
  geom_smooth(se = FALSE)
```

```
# color aesthetic specified for only the geom_point layer
ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point(aes(color = class)) +
  geom_smooth(se = FALSE)
```

https://uc-r.github.io/ggplot_intro

Why use ggplot2?

week one ....



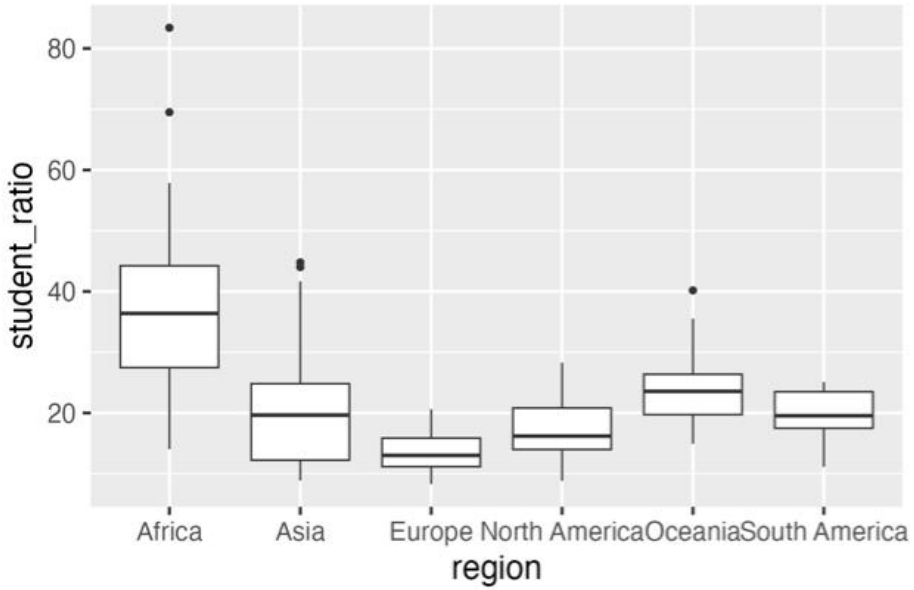
* figure is totally fabricated but, I claim, s

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

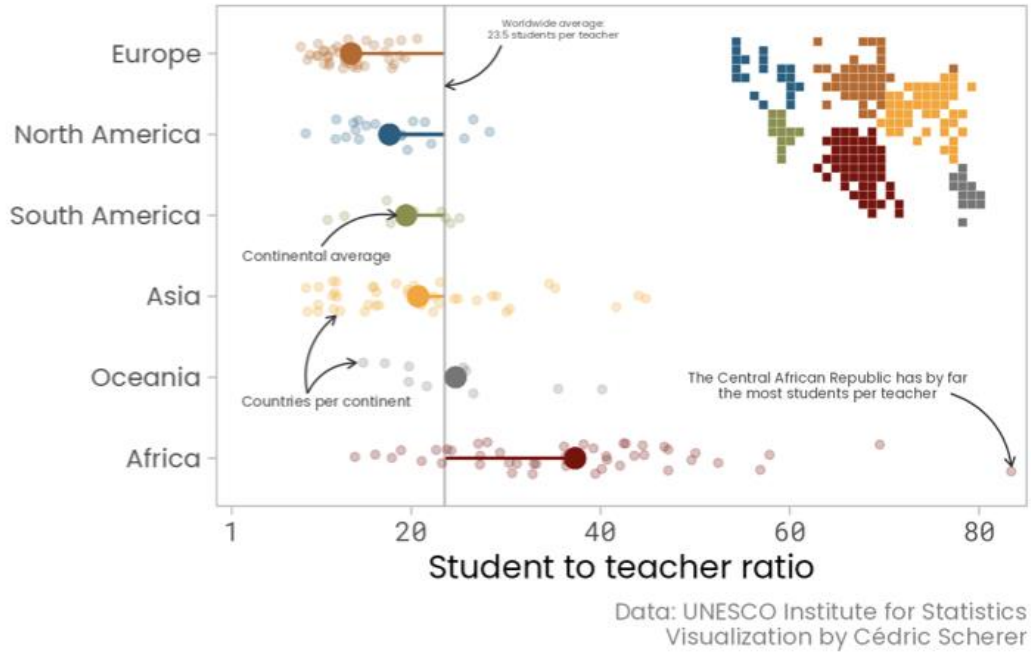
* figure is totally fabricated but, I claim, still true

Scales scale_.
Statistical transformations stat_:
Coordinate system coord_:
Facets facet_:
Visual themes theme():

Before

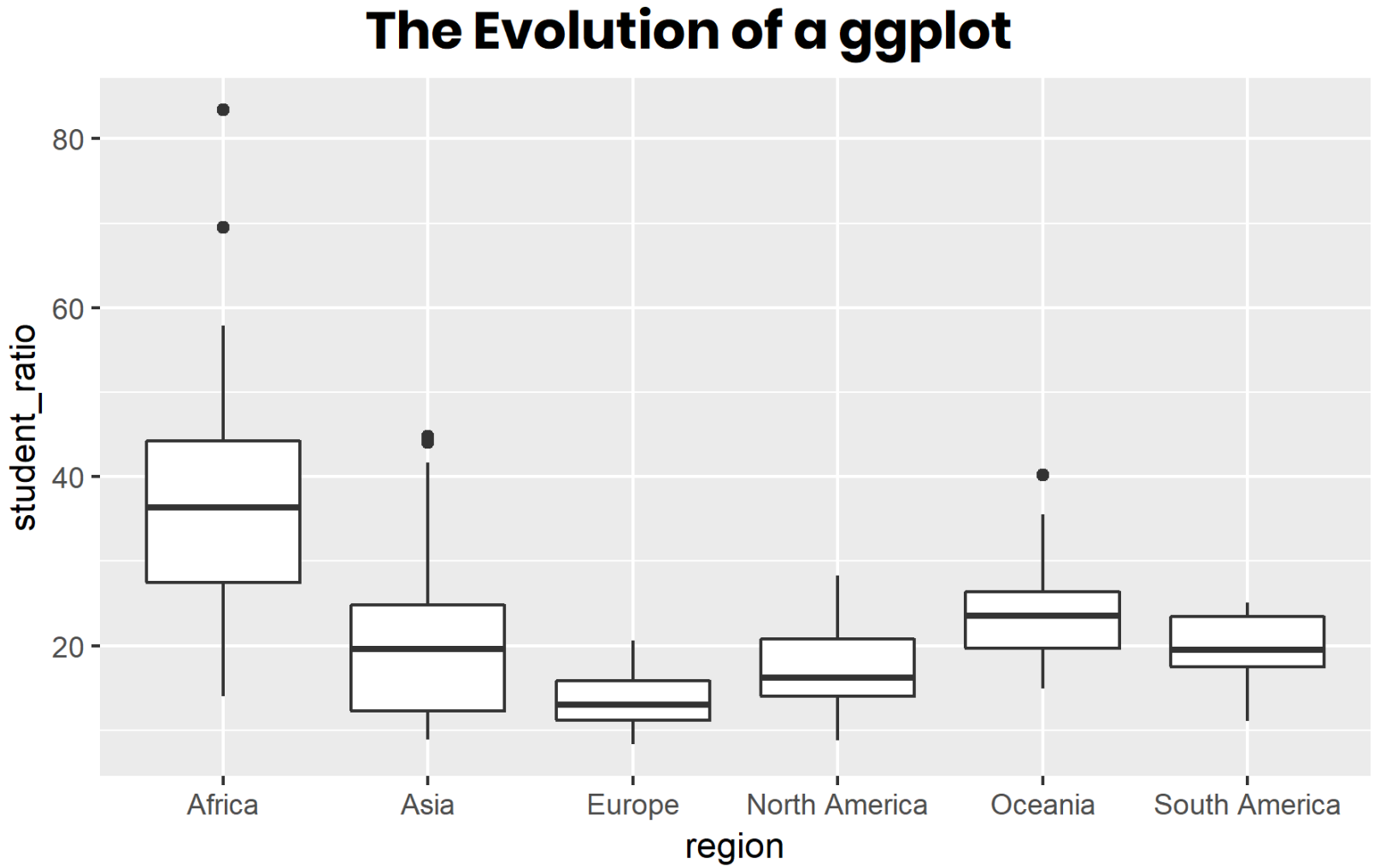


After



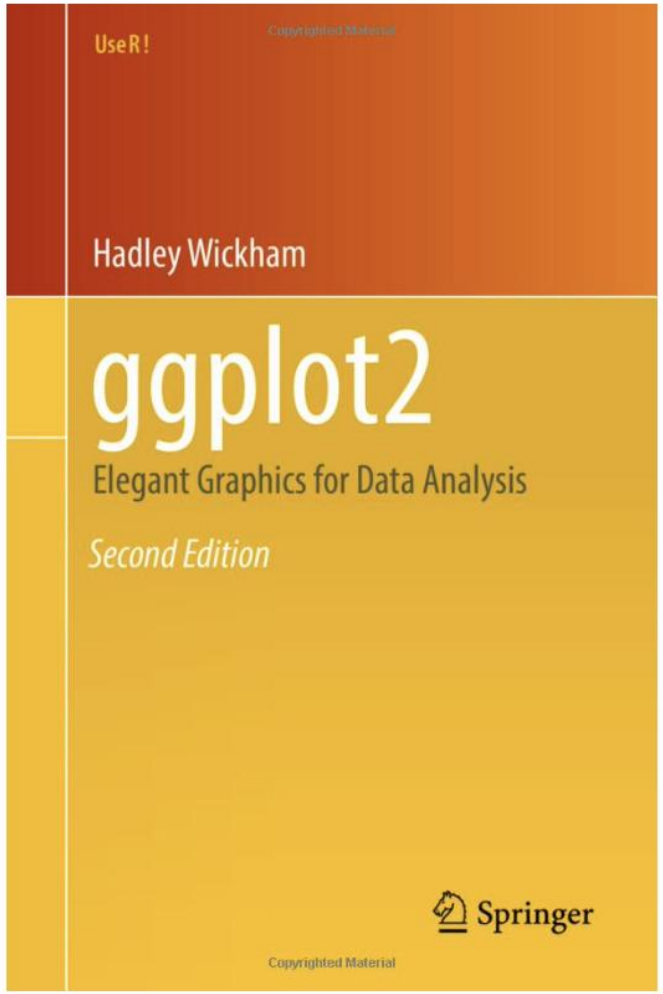
https://cedricscherer.netlify.app/2019/05/17/the-evolution-of-a-ggplot/

How does it work? Simple and powerful



https://cedricscherer.netlify.app/2019/05/17/the-evolution-of-a-ggplot/

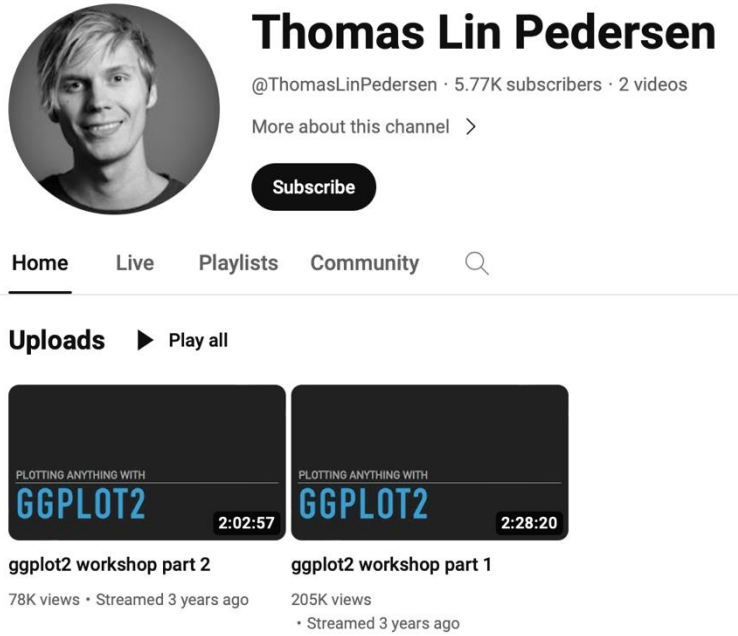
Learn more! Power in community



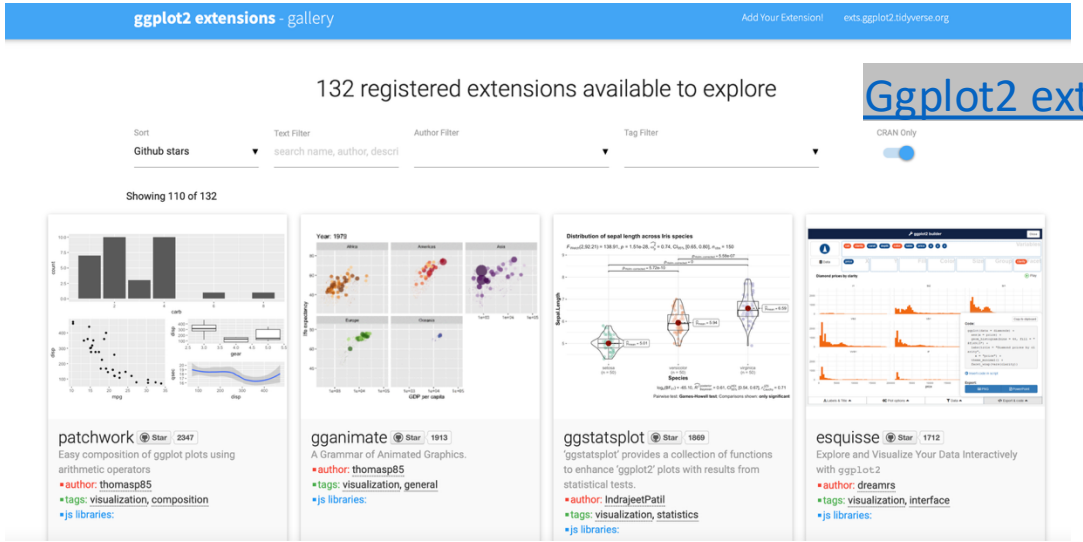
https://ggplot2-book.org

Core developers of ggplot2 are:

- Winston Chang
- Lionel Henry
- Thomas Lin Pedersen
- Kohske Takahashi
- Claus Wilke
- Kara Woo
- Hiroaki Yutani
- Dewey Dunnington



Youtube ggplot2 tutorial



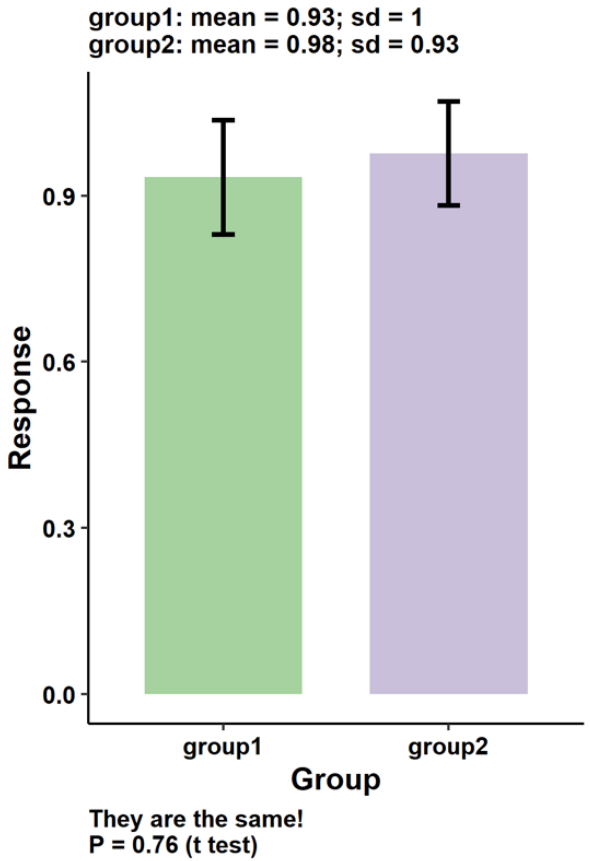
Ggplot2 extensions

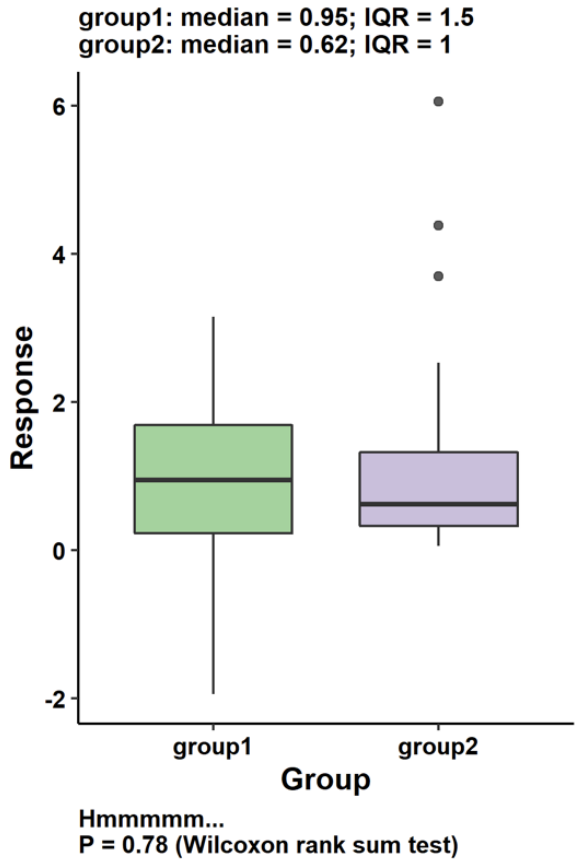
Awesome resources

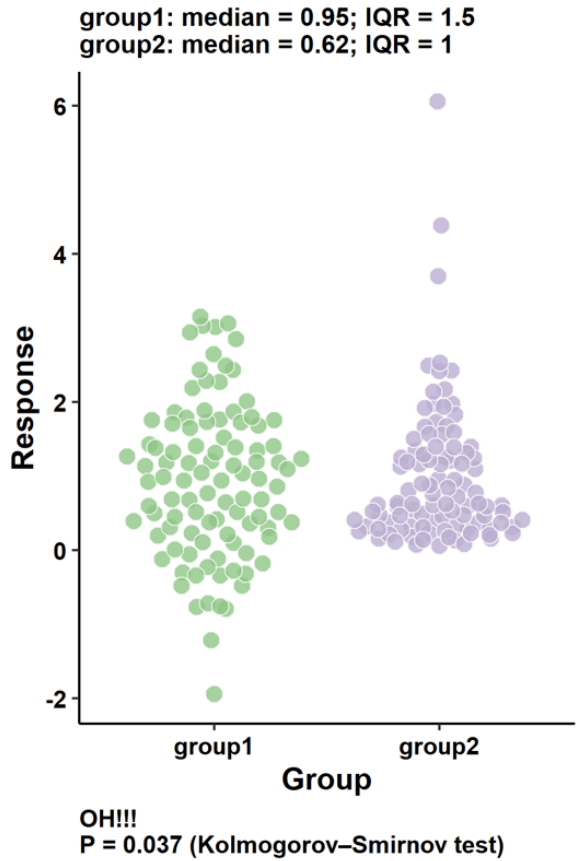
- https://github.com/erikgahner/awesome-ggplot2
- https://r-graph-gallery.com/index.html
- http://www.sthda.com/english/wiki/ggplot2-essentials
- Very good tricks: https://github.com/teunbrand/ggplot_tricks, https://albert-rapp.de/posts/ggplot2-tips/21_common_ggplot_mistakes/21_common_ggplot_mistakes.html
- https://r-charts.com
- https://waterdata.usgs.gov/blog/ggplot-jazz/
- https://www.cedricscherer.com/top/dataviz/
- https://cedricscherer.netlify.app/2019/08/05/a-ggplot2-tutorial-for-beautiful-plotting-in-r/

Advice on choosing or creating plots

Mean separation plots and limitations of bar graphs

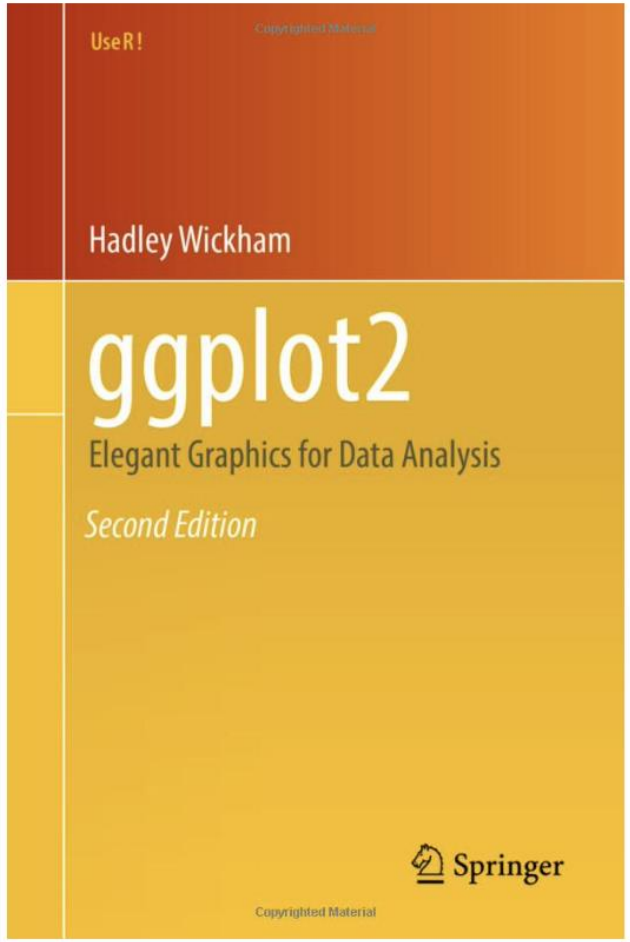






https://github.com/cxli233/FriendsDontLetFriends

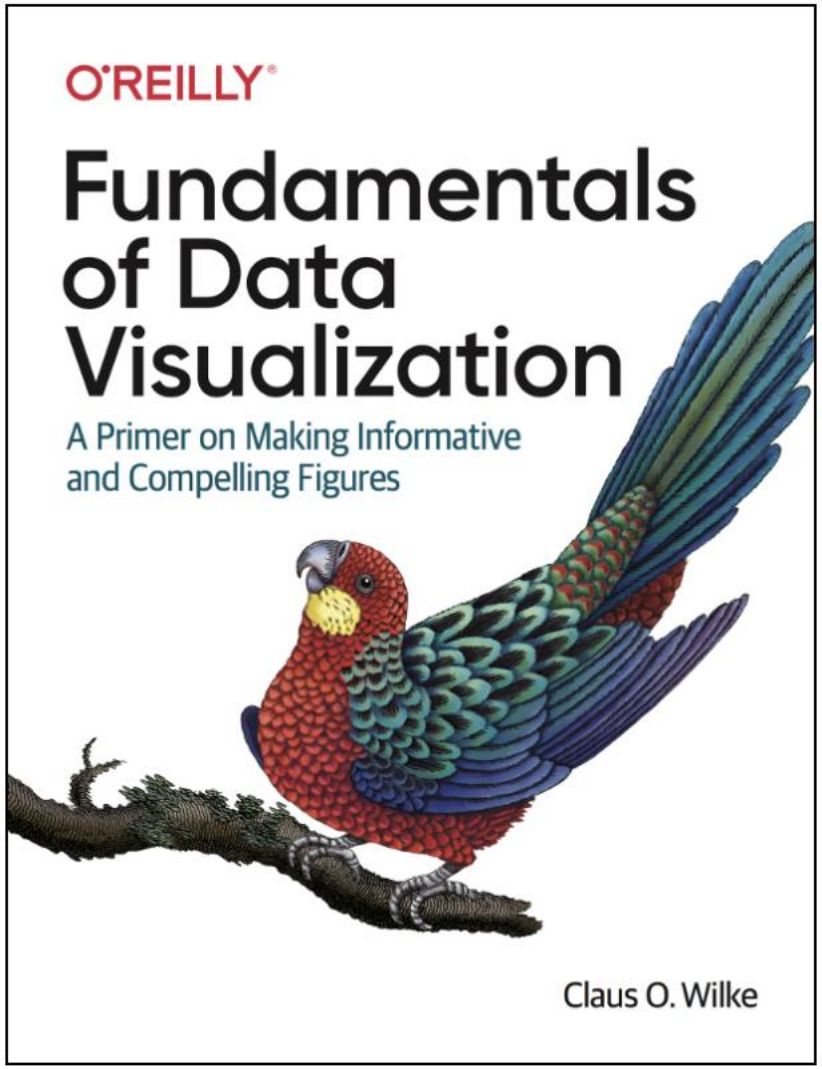
Learn more! Power in community



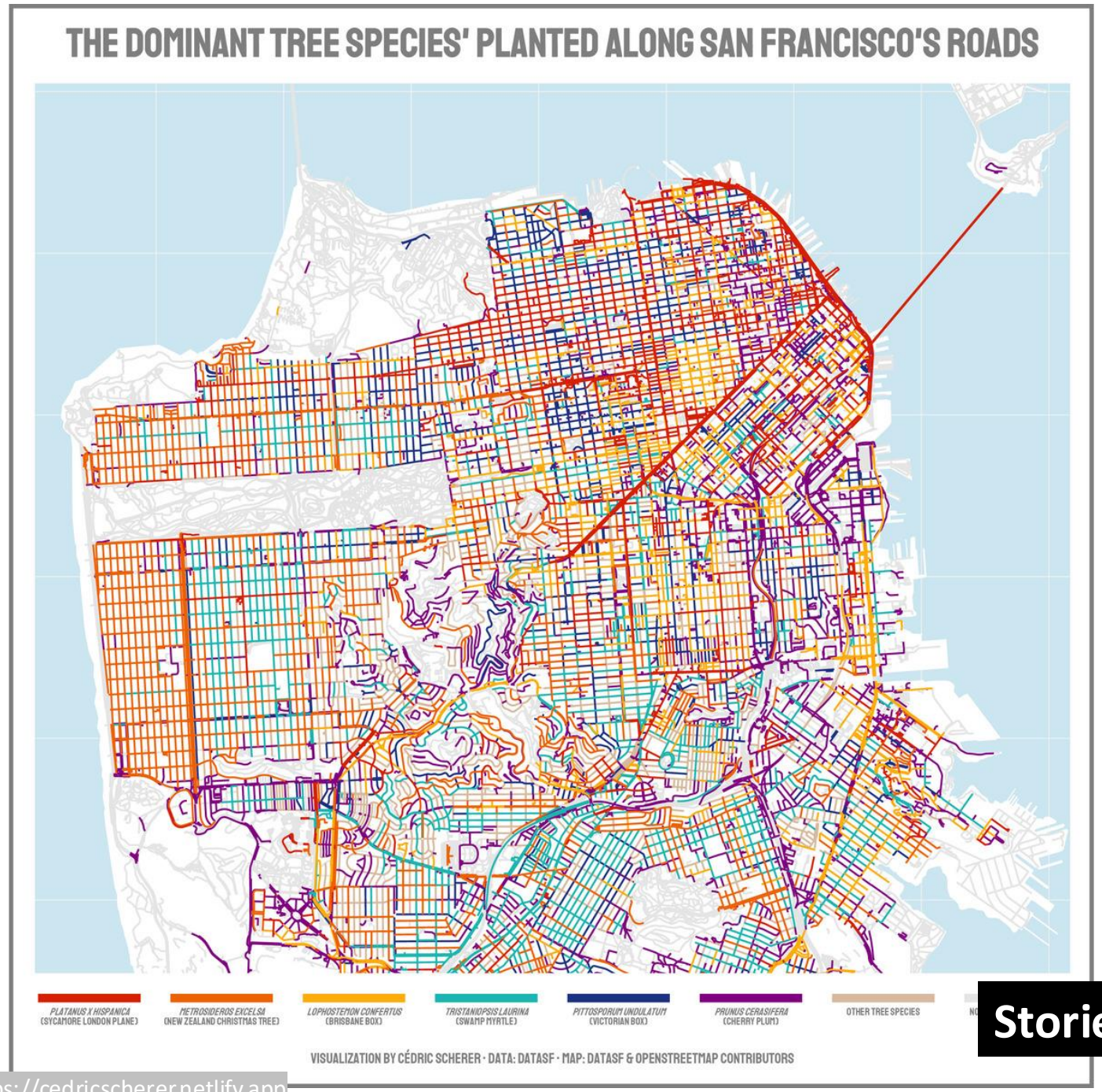
https://ggplot2-book.org

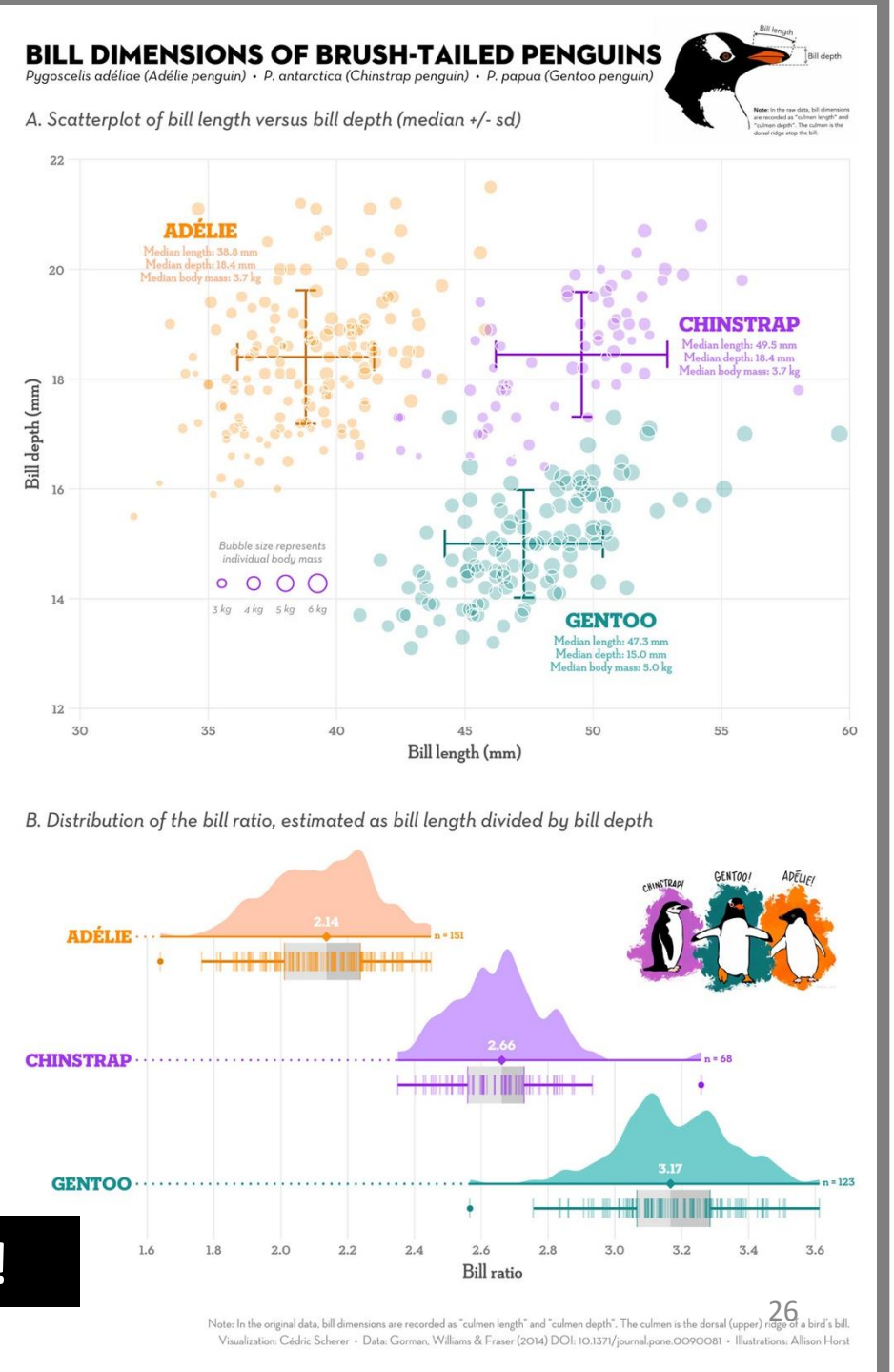
Core developers of ggplot2 are:

- Winston Chang
- Lionel Henry
- Thomas Lin Pedersen
- Kohske Takahashi
- Claus Wilke
- Kara Woo
- Hiroaki Yutani
- Dewey Dunnington

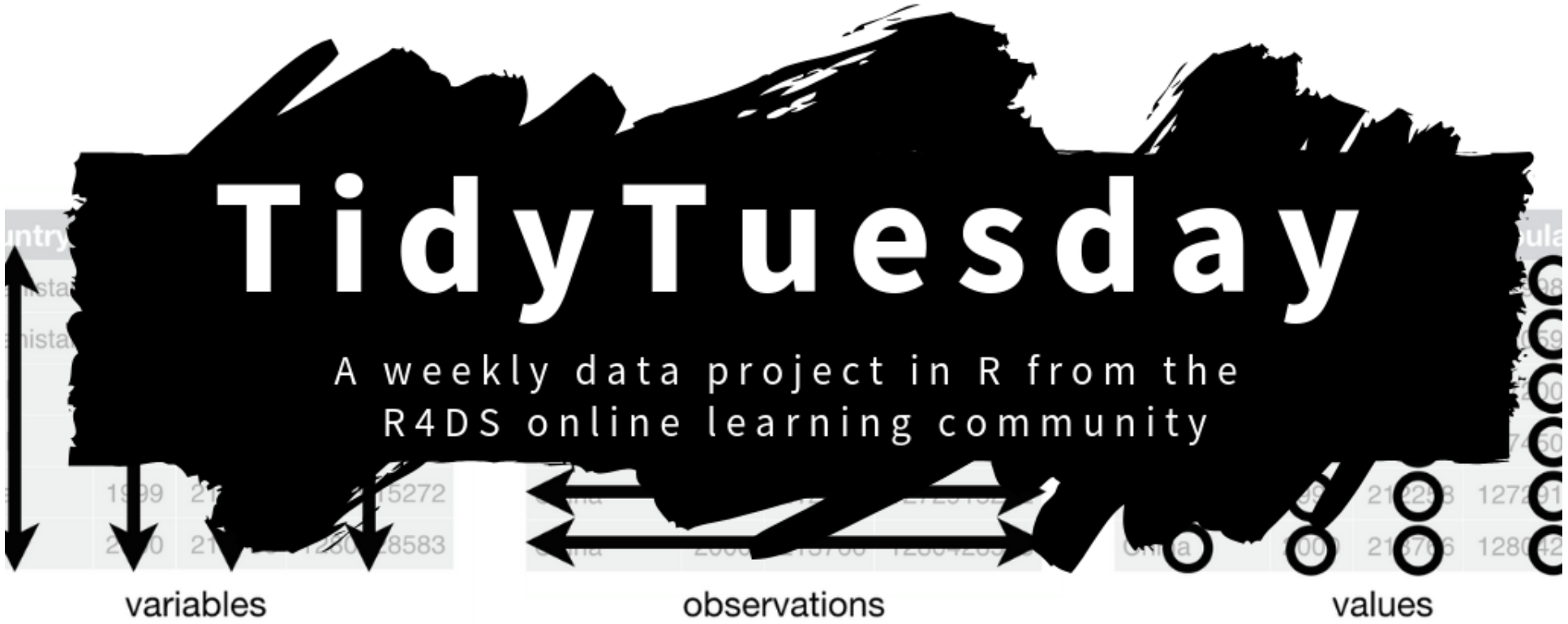


https://clauswilke.com/dataviz/


THE DOMINANT TREE SPECIES' PLANTED ALONG SAN FRANCISCO'S ROADS
PLATANUS X HISPANICA (SYCAMORE LONDON PLANE)
METROSIDEROS EXCELSA (NEW ZEALAND CHRISTMAS TREE)
LOPHOSTEMON CONFERTUS (BRISBANE BOX)
TRISTANIOPSIS LAURINA (SWAMP MYRTLE)
PITTOSPORUM UNDULATUM (VICTORIAN BOX)
PRUNUS CERASIFERA (CHERRY PLUM)
OTHER TREE SPECIES
VISUALIZATION BY CÉDRIC SCHERER · DATA: DATASF · MAP: DATASF & OPENSTREETMAP CONTRIBUTORS


BILL DIMENSIONS OF BRUSH-TAILED PENGUINS
Pygoscelis adéliae (Adélie penguin) • P. antarctica (Chinstrap penguin) • P. papua (Gentoo penguin)
Bill length
Bill depth
A. Scatterplot of bill length versus bill depth (median +/- sd)
ADÉLIE
Median length: 38.8 mm
Median depth: 18.4 mm
Median body mass: 3.7 kg
CHINSTRAP
Median length: 49.5 mm
Median depth: 18.4 mm
Median body mass: 3.7 kg
GENTOO
Median length: 47.3 mm
Median depth: 15.0 mm
Median body mass: 5.0 kg
Bubble size represents individual body mass
3 kg 4 kg 5 kg 6 kg
Bill depth (mm)
Bill length (mm)
B. Distribution of the bill ratio, estimated as bill length divided by bill depth
2.14
n = 151
2.66
n = 68
3.17
n = 123
Bill ratio
Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0090081 • Illustrations: Allison Horst

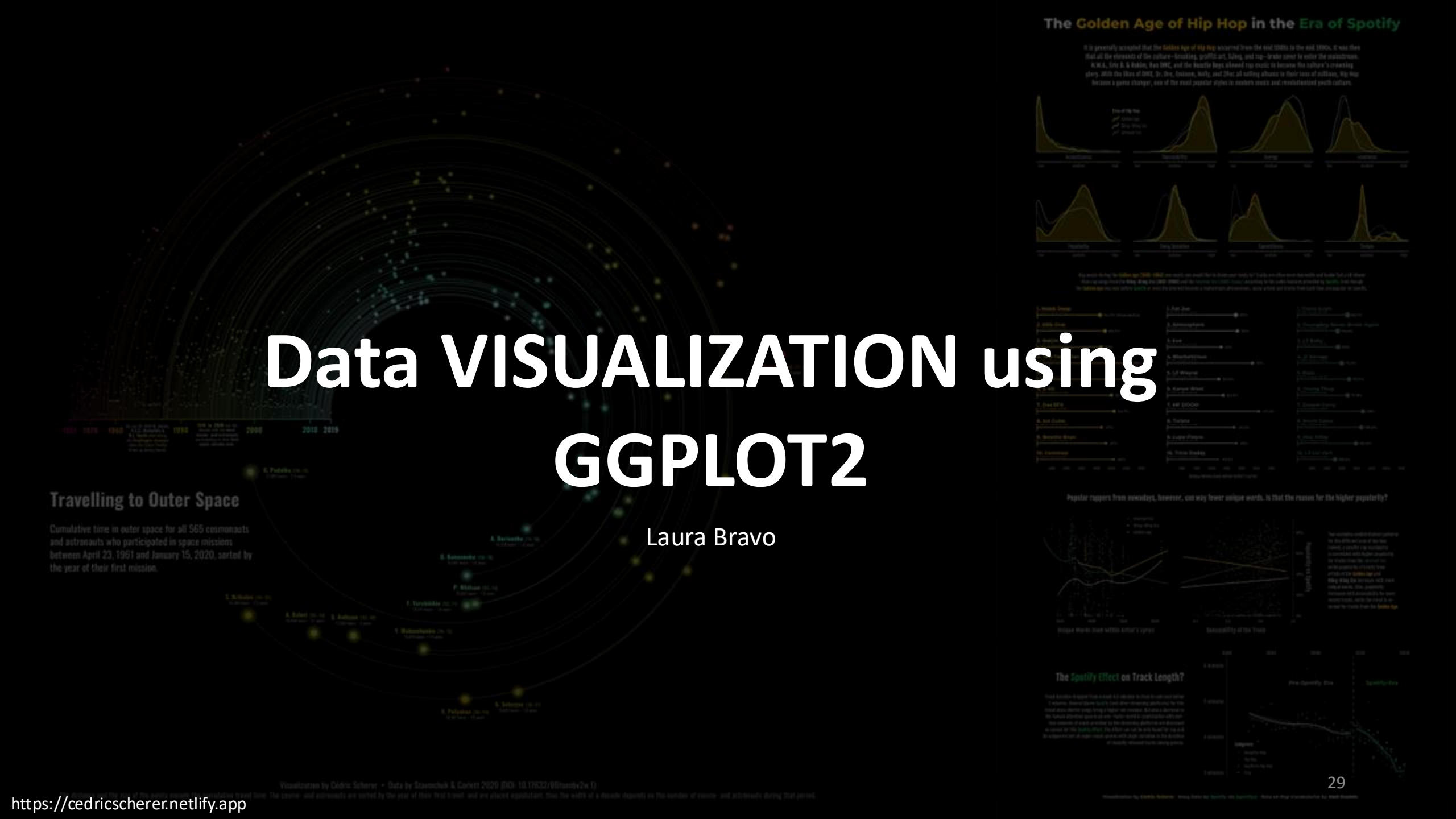
Stories!


TidyTuesday
A weekly data project in R from the
R4DS online learning community
variables
observations
values

Join the R User Group!

- Join our mailing list:

  Email Majordomo@lists.bham.ac.uk with **subscribe uobrug** in the main body of a plain text email (please include a new line at the end and no other content, e.g. signatures).



Data VISUALIZATION using GGPLOT2

Laura Bravo

https://cedricscherer.netlify.app