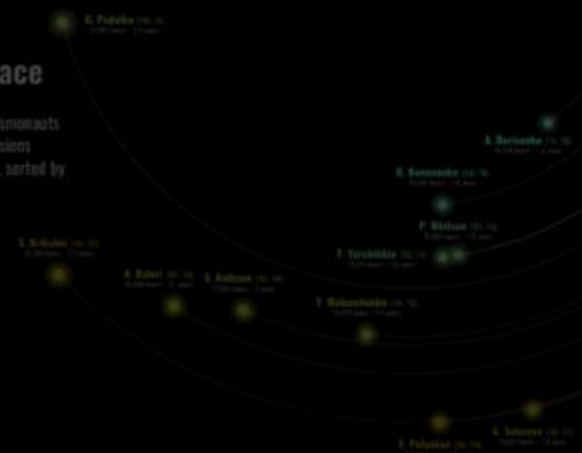


Data VISUALIZATION using GGPLOT2

Laura Bravo

Travelling to Outer Space

Cumulative time in outer space for all 565 cosmonauts and astronauts who participated in space missions between April 23, 1961 and January 15, 2020, sorted by the year of their first mission.

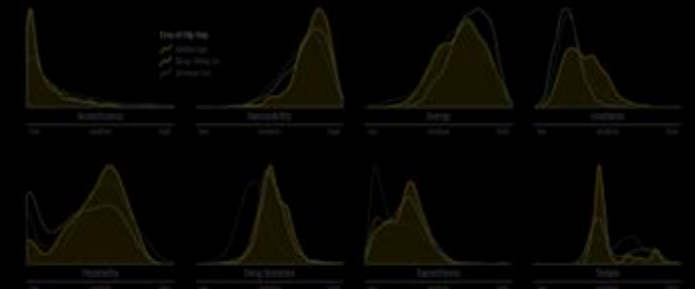


Visualization by Cedric Scherer - Data by Stawickich & Cornett 2020 (DOI: 10.17632/85tmmv2w1)

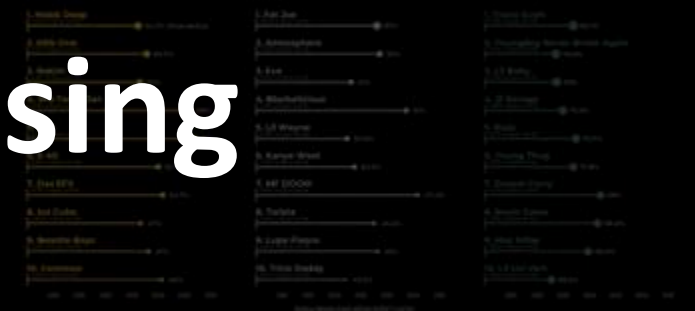
The cosine- and astronomical data are sorted by the year of their first mission, and are placed equidistant. Thus, the width of a decade depends on the number of cosine- and astronomical data during that period.

The Golden Age of Hip Hop in the Era of Spotify

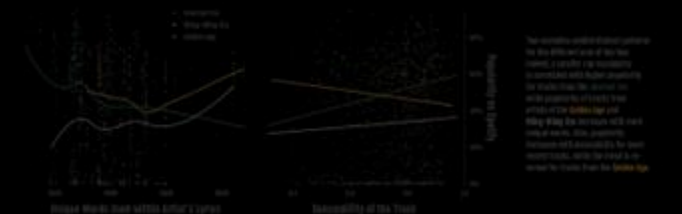
It is generally accepted that the **Golden Age of Hip Hop** occurred from the mid 1980s to the mid 1990s. It was then that all the elements of the culture – freestyle, graffiti art, b-boying, and rap – broke cover to enter the mainstream. K.R.A., Eric B. & Rakim, Run DMC, and the Beastie Boys cleaned rap music to become the culture's crowning glory. With the likes of DMX, Dr. Dre, Jay-Z, Nas, and 50 Cent all selling albums to their tens of millions, Hip Hop became a genre changer, one of the most popular styles in modern music and reinvigorated youth culture.



Why were things like **Golden Age (1980-1990)** and **Old School (1970-1980)** not more successful than other eras? The answer lies in the fact that these eras were not as well-timed as the **Golden Age** and **Old School** eras. The **Golden Age** and **Old School** eras were the only eras that were able to capture the mainstream market, while other eras were too niche or too early to do so.

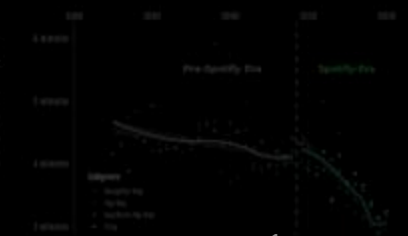


Popular rappers from nowadays, however, can say fewer unique words. Is that the reason for the higher popularity?



The Spotify Effect on Track Length?

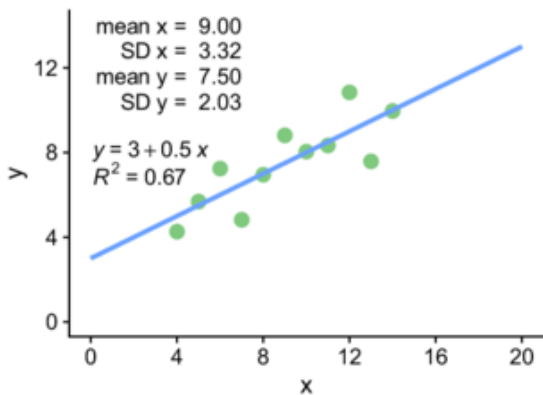
Track lengths in songs have increased over time, but is this due to the Spotify effect? The answer lies in the fact that Spotify has a higher average track length than other streaming services. This is because Spotify's algorithm favors longer tracks, which are more likely to be played for a longer duration. This has led to an increase in the average track length of songs released on Spotify.



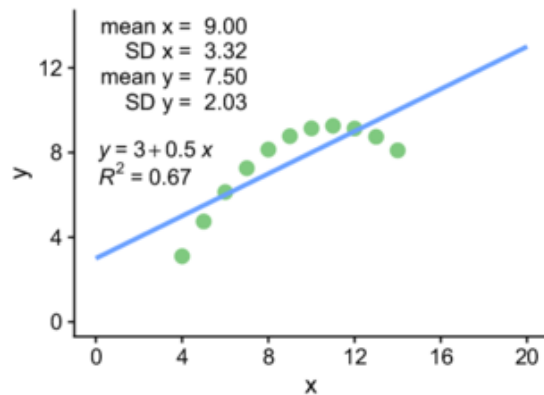
Always, always, always plot the data

Anscombe's quartet (Anscombe, 1973)

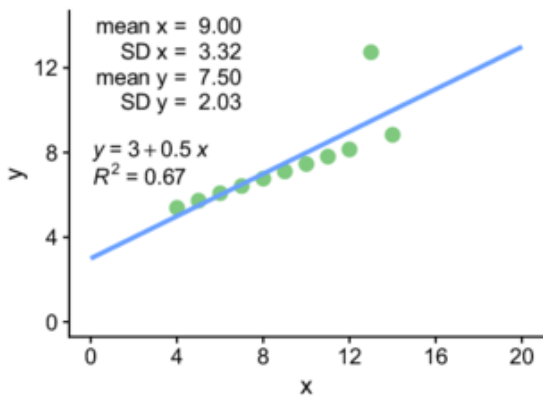
Dataset 1



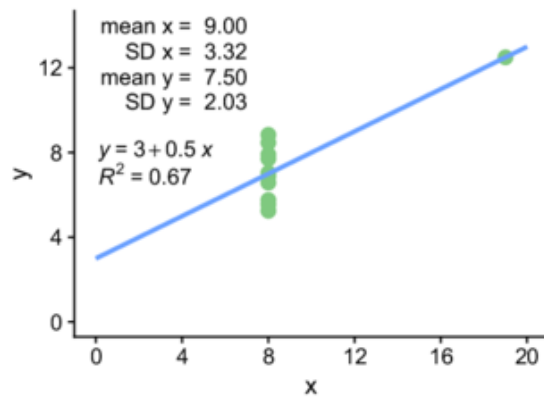
Dataset 2



Dataset 3

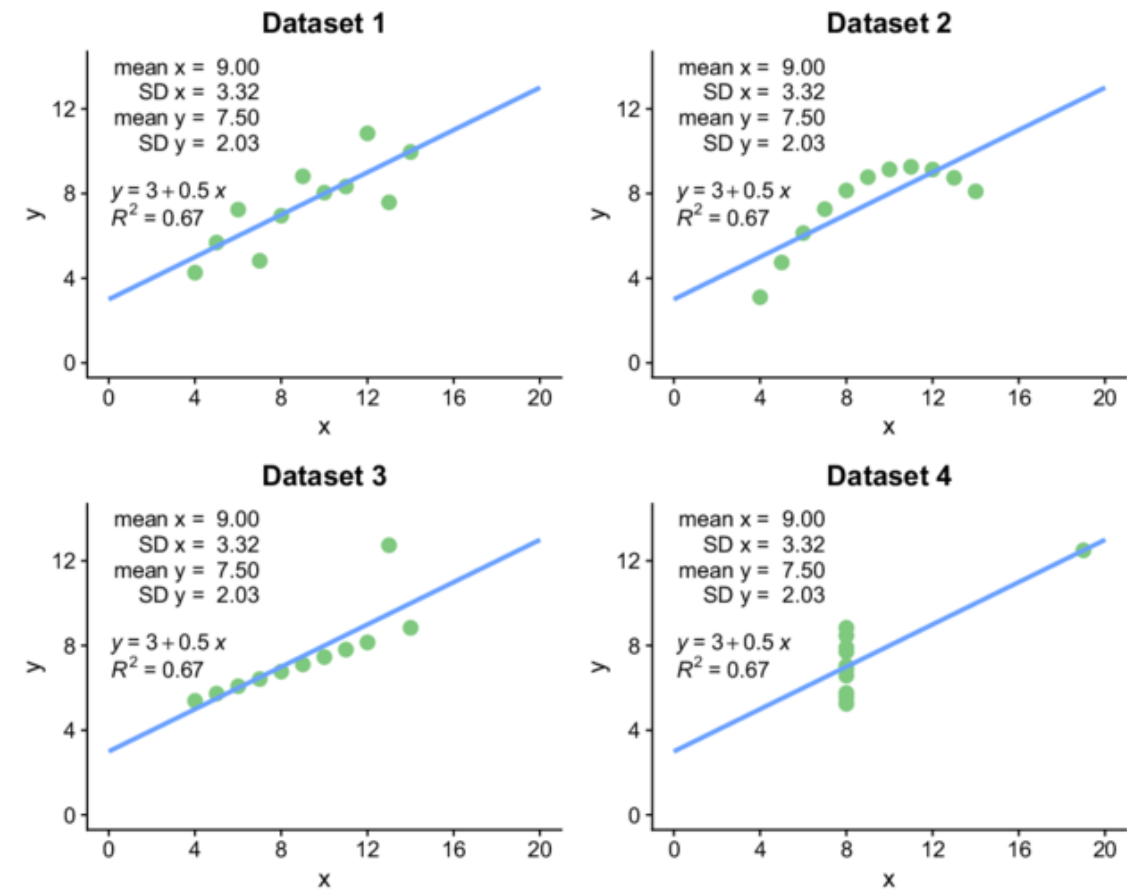


Dataset 4

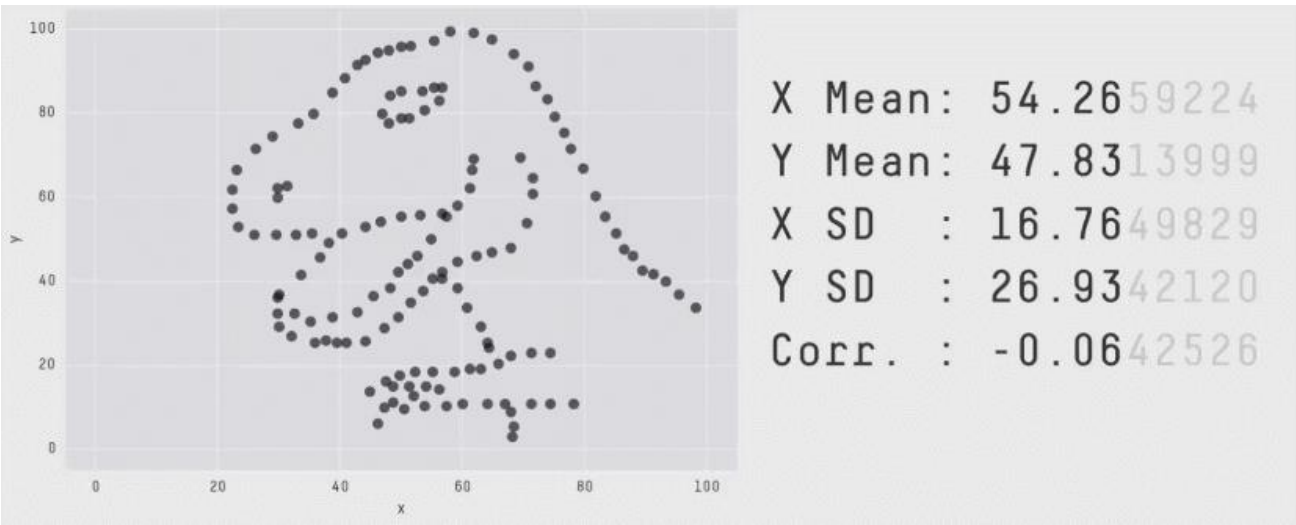


Always, always, always plot the data

Anscombe's quartet (Anscombe, 1973)



Datasaurus dozen (Matejka & Fitzmaurice, 2017)



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

Graphics in R

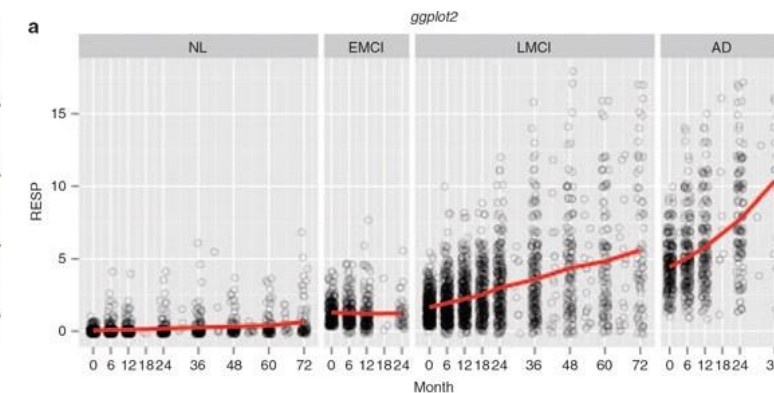
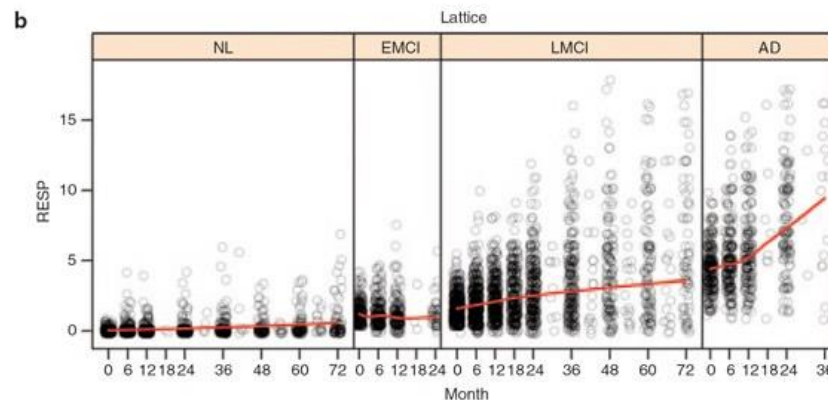
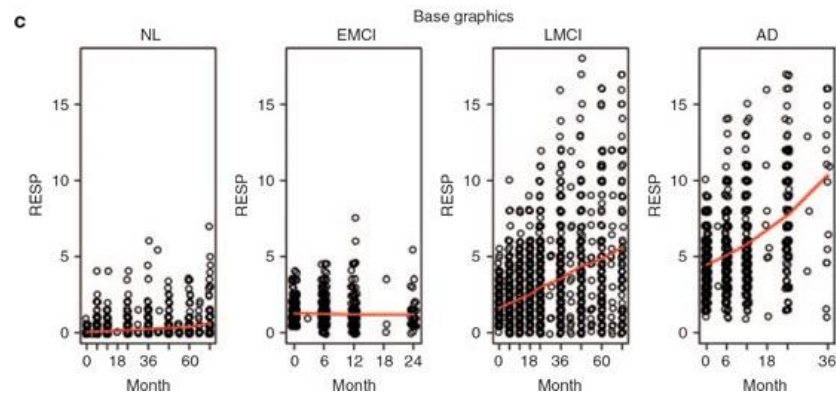
Base R (1993)

Lattice (1993) (2008)

Trellis graphics – Bell Laboratories

Ggplot2 (2005)

Layered grammar of graphics

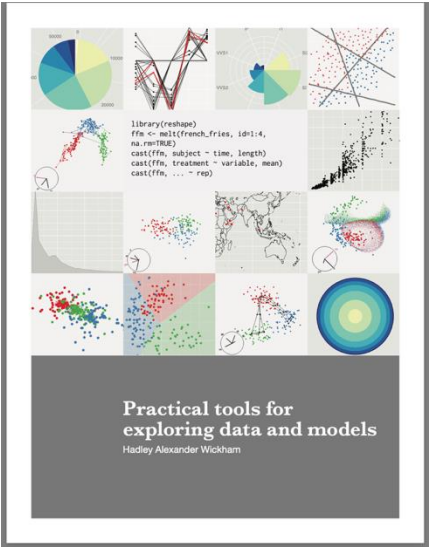
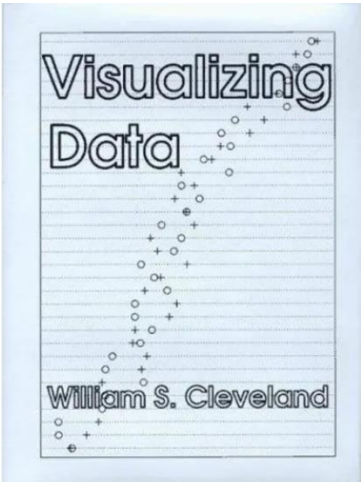
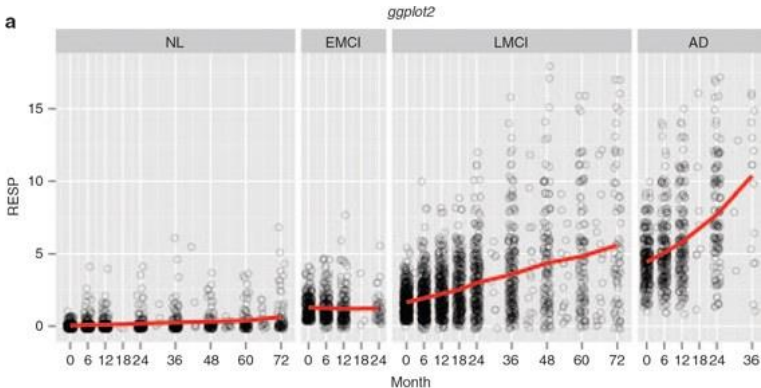
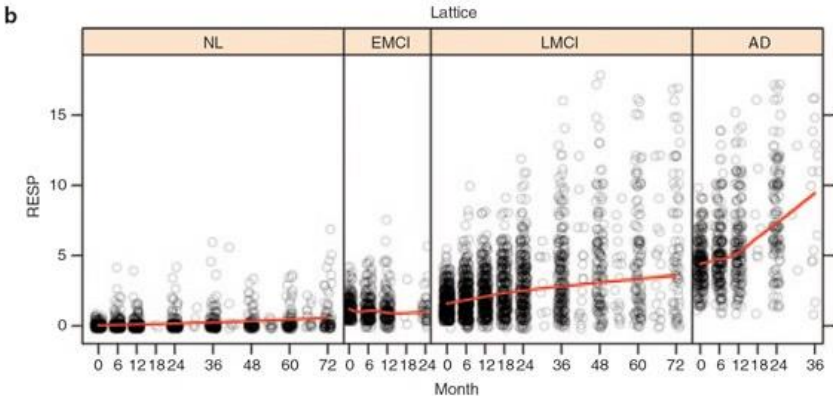
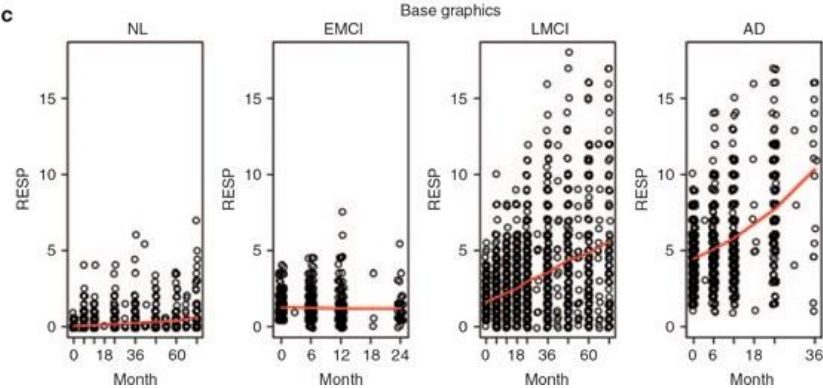


Graphics in R

Base R (1993)

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

Ggplot2 (2005)
Layered grammar of graphics

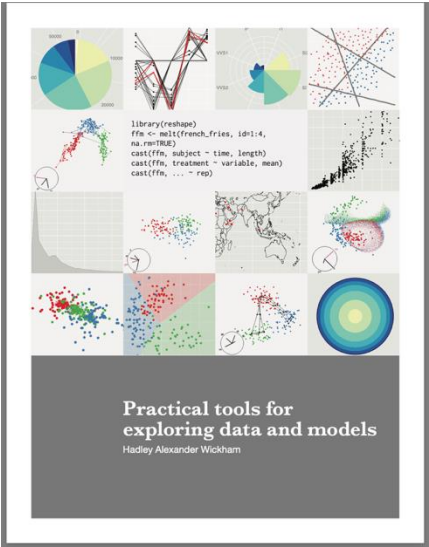
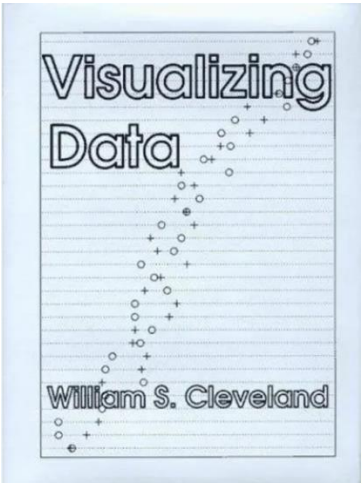
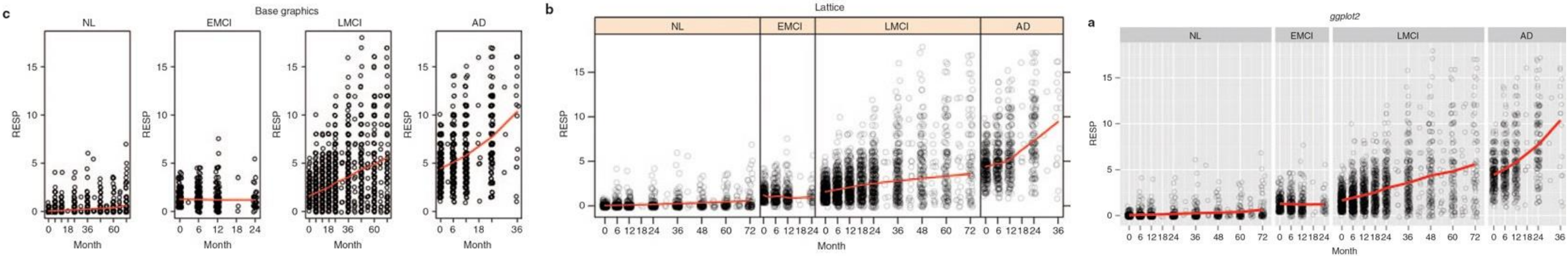


Graphics in R

Base R (1993)

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

Ggplot2 (2005)
Layered grammar of graphics

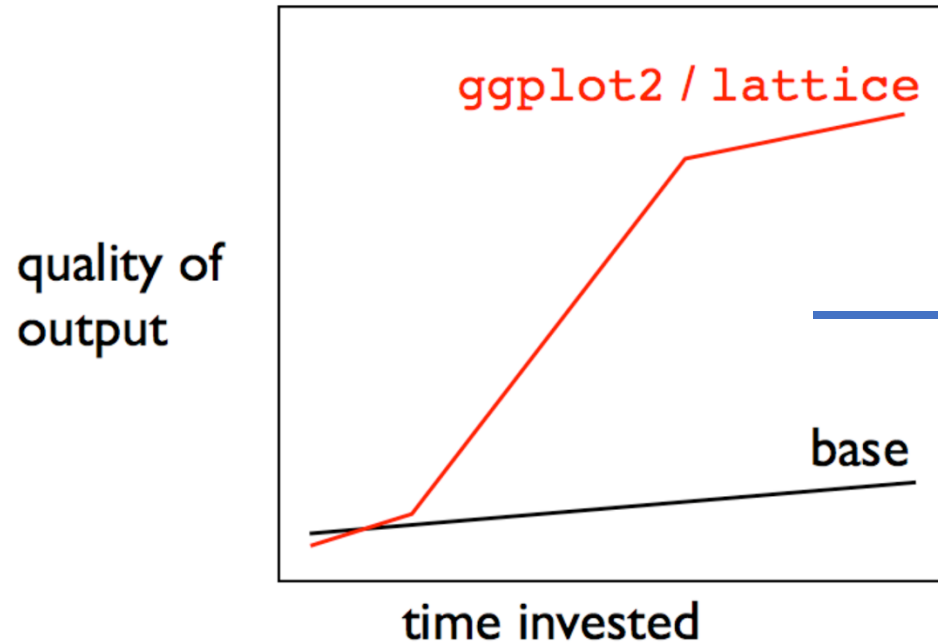


Application of ggplot2 to Pharmacometric Graphics (Kaori Ito, 2013)

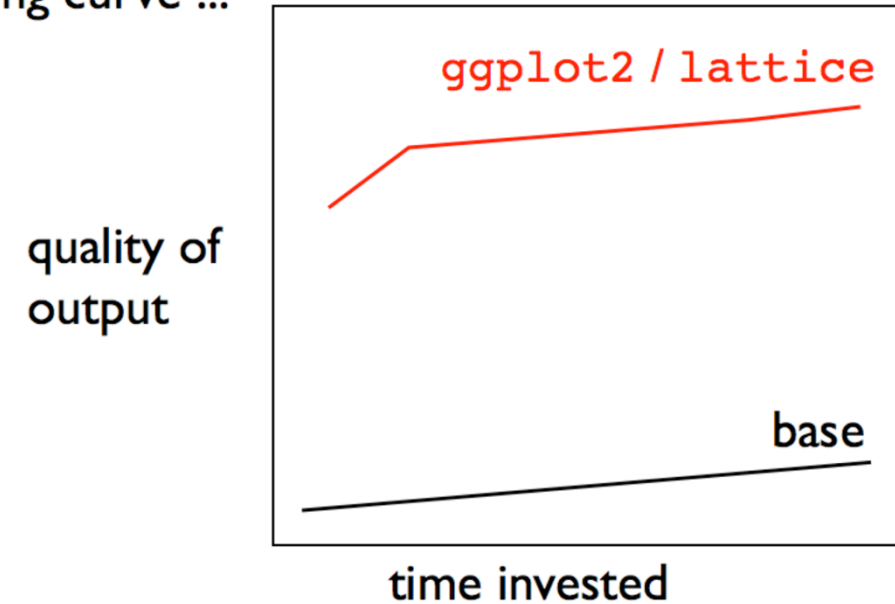
Hadley Wickham

Why use ggplot2?

week one



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



* figure is totally fabricated but, I claim, :

* 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

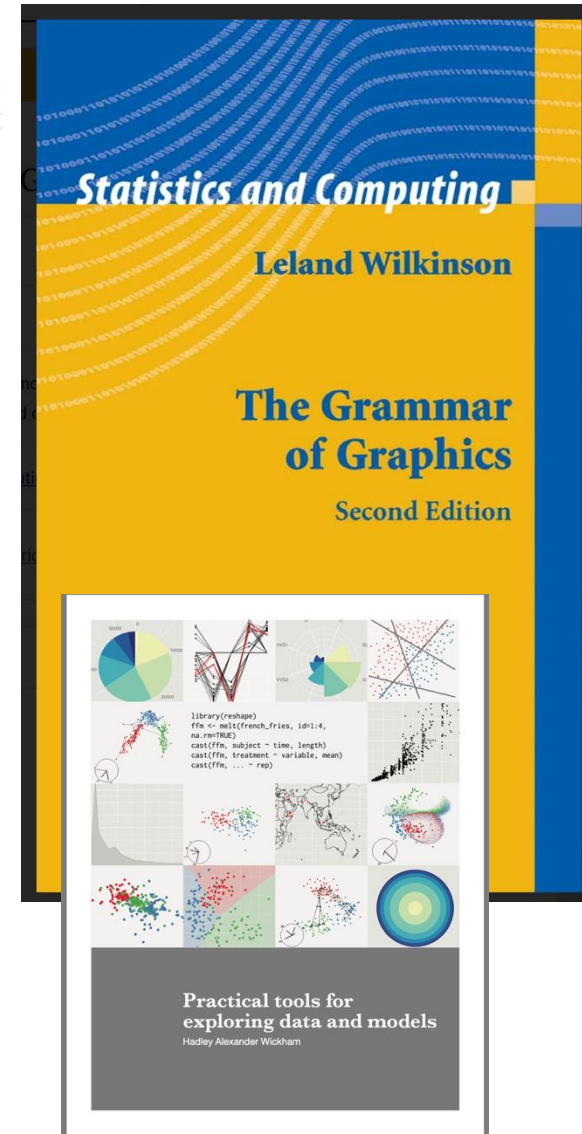
Data: The raw data that you want to plot.

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_`:

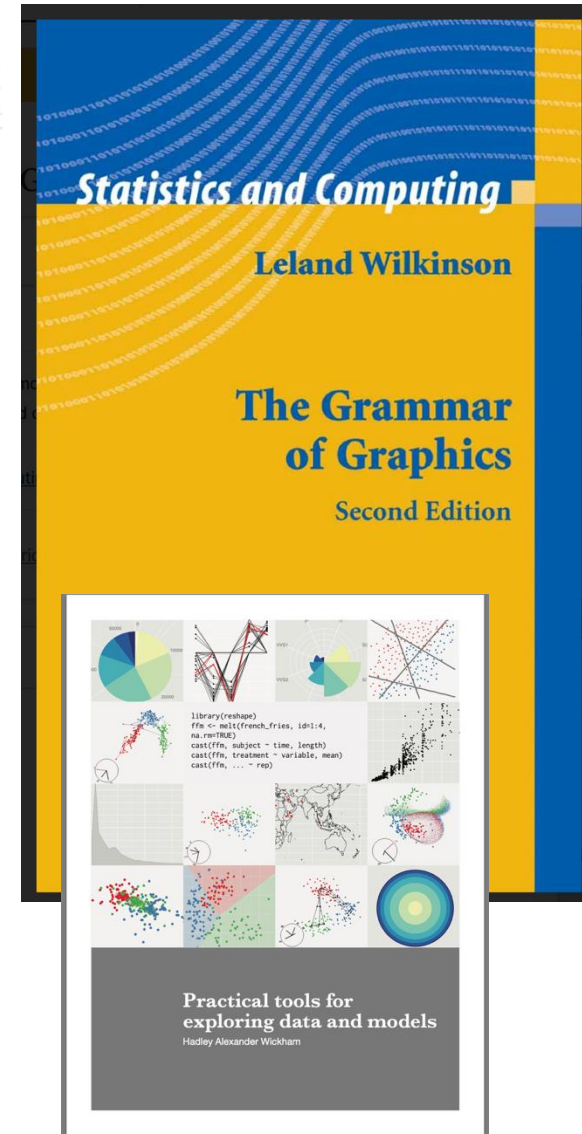
Coordinate system `coord_`:

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:

```
ggplot (data = <DATA>) +  
  <GEOM_FUNCTION> (mapping = aes(<MAPPINGS>),  
    stat = <STAT> , position = <POSITION> ) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

required

Not required, sensible defaults supplied

ggplot(data = mpg, **aes**(x = cty, y = hwy)) Begins a plot that you finish by adding layers to. Add one geom function per layer.

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>
## 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>
```

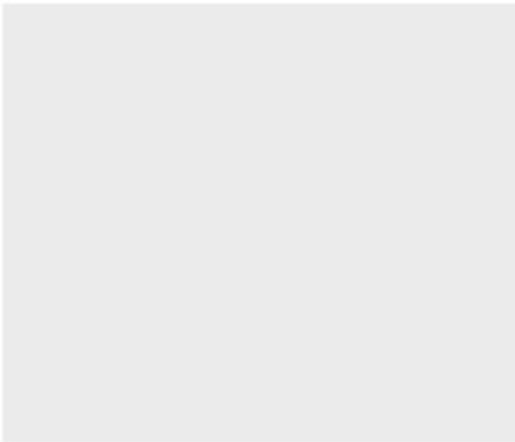
How does it work?

```
# create canvas
ggplot(mpg)

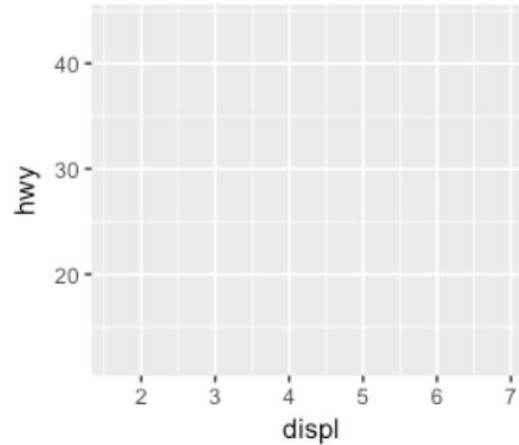
# variables of interest mapped
ggplot(mpg, aes(x = displ, y = hwy))

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

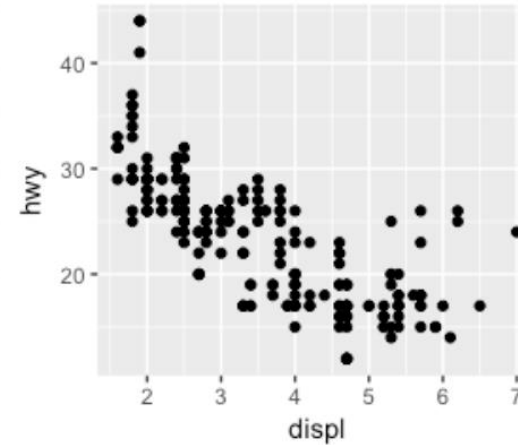
Canvas



Canvas + variables mapped to axes



Data plotted

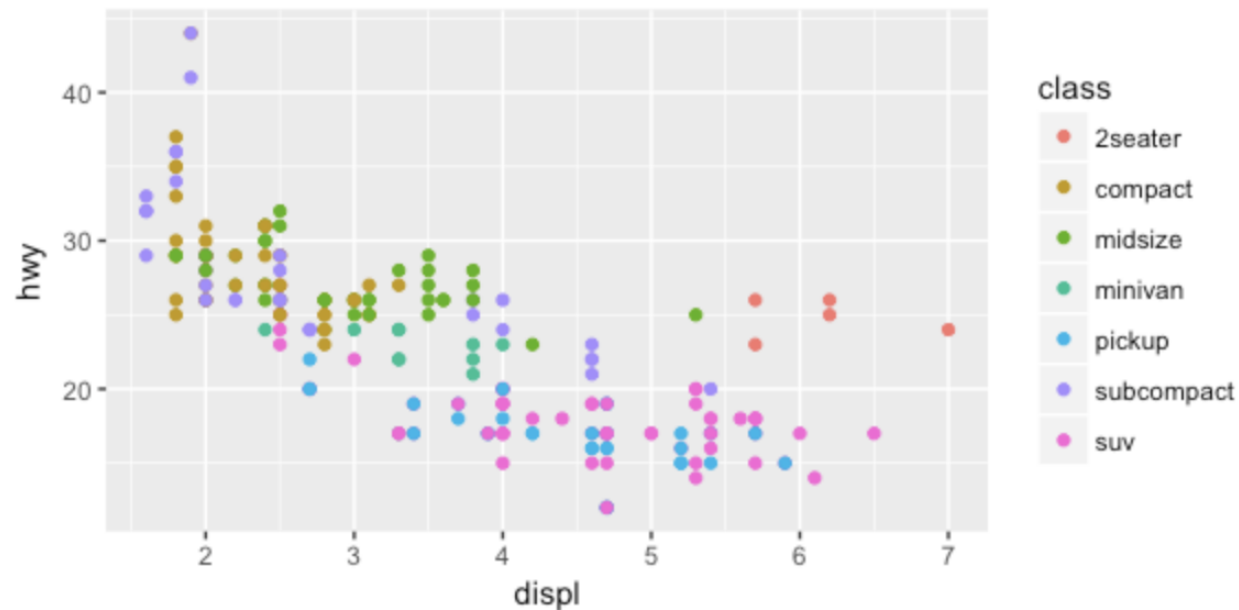


Layers!!!

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()
```



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.

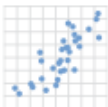
TWO VARIABLES

continuous x , continuous y

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



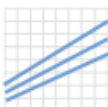
e + geom_label() (aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE) x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust



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



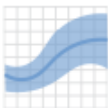
e + geom_point() x, y, alpha, color, fill, shape, size, stroke



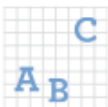
e + geom_quantile() x, y, alpha, color, group, linetype, size, weight



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



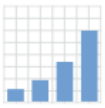
e + geom_smooth() (method = lm) x, y, alpha, color, fill, group, linetype, size, weight



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

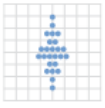
```
f <- ggplot(mpg, aes(class, hwy))
```



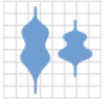
f + geom_col() x, y, alpha, color, fill, group, linetype, size



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

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

Fundamental building blocks

How does it work? Geometric shapes

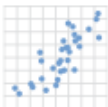
TWO VARIABLES

continuous x , continuous y

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



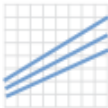
e + geom_label() aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust



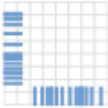
e + geom_jitter() height = 2, width = 2, x, y, alpha, color, fill, shape, size



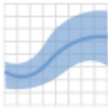
e + geom_point(), x, y, alpha, color, fill, shape, size, stroke



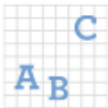
e + geom_quantile(), x, y, alpha, color, group, linetype, size, weight



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



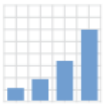
e + geom_text() aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE, x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

scatterplot

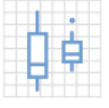
barchart

discrete x , continuous y

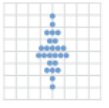
```
f <- ggplot(mpg, aes(class, hwy))
```



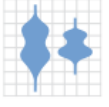
f + geom_col(), x, y, alpha, color, fill, group, linetype, size



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

Box-and-whisker plot

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

Fundamental building blocks

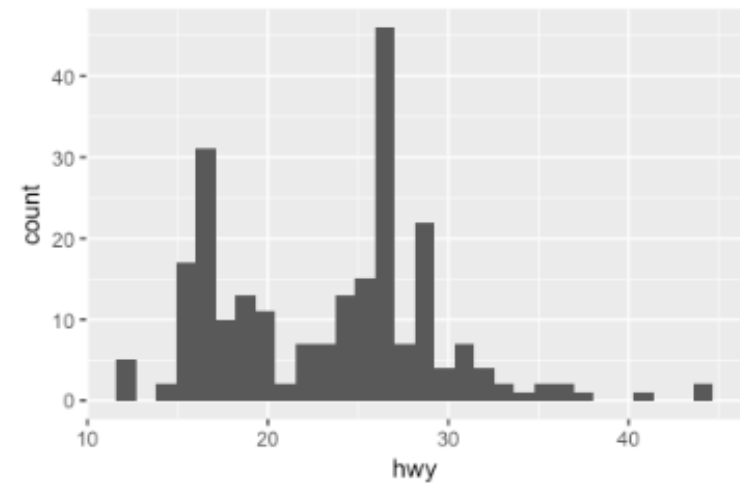
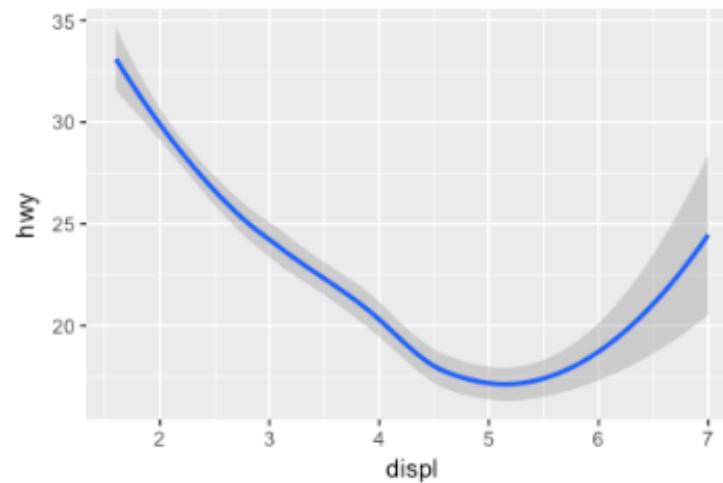
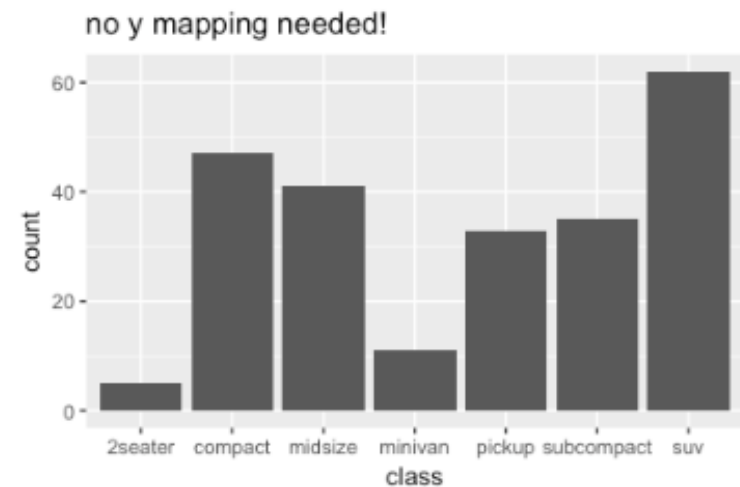
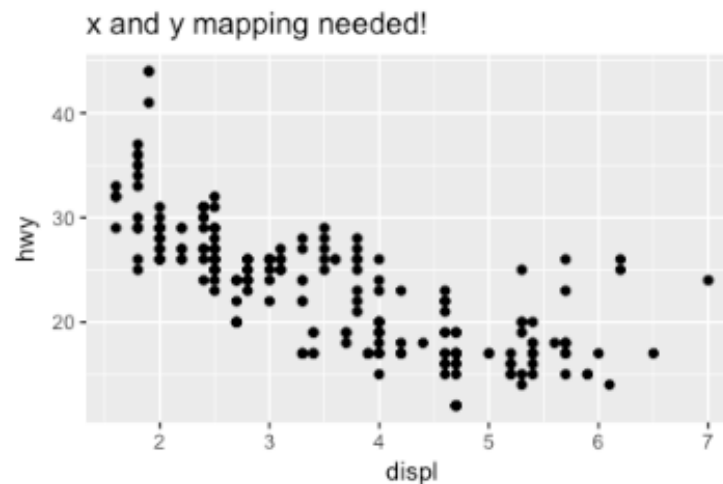
How does it work? Geometric shapes

```
# 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()
```

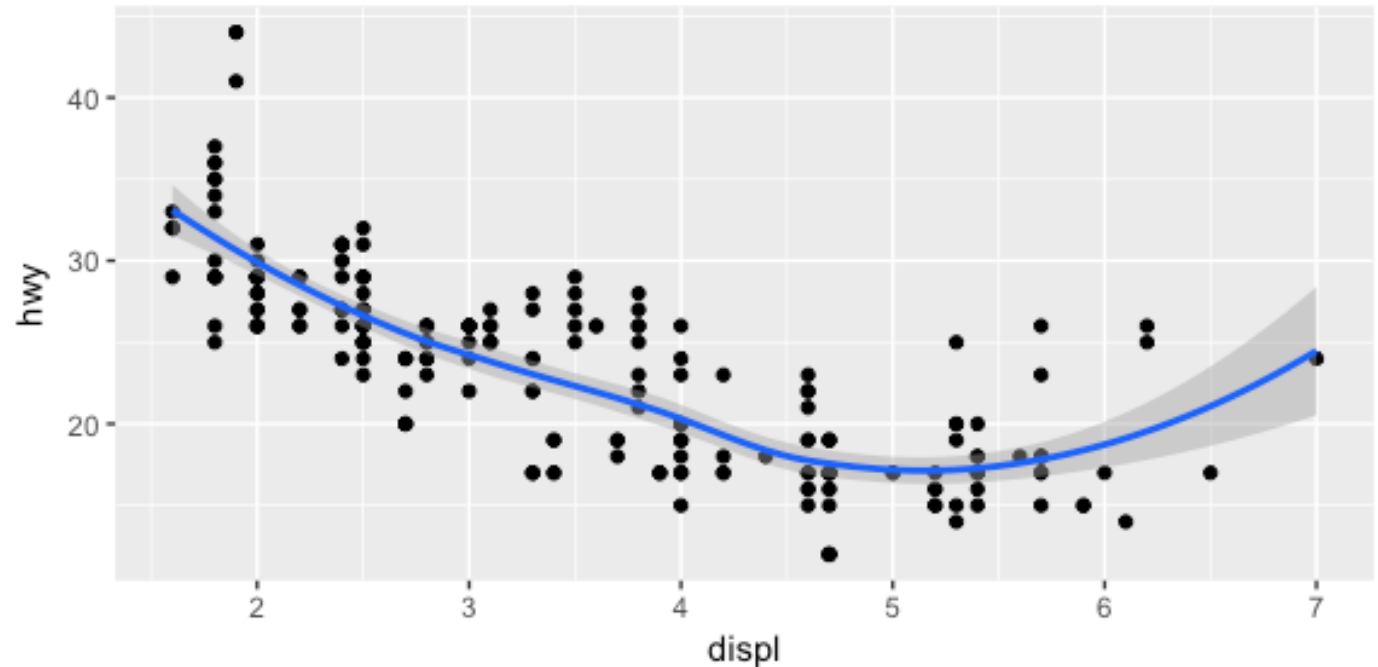


How does it work? Geometric shapes

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.

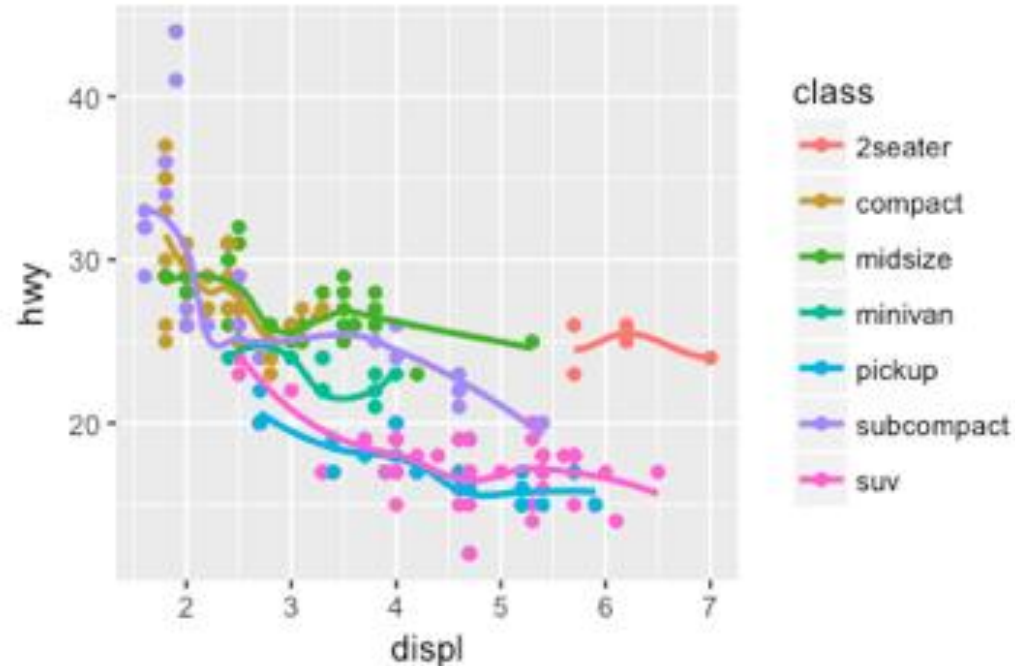
Each function returns a layer!!

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

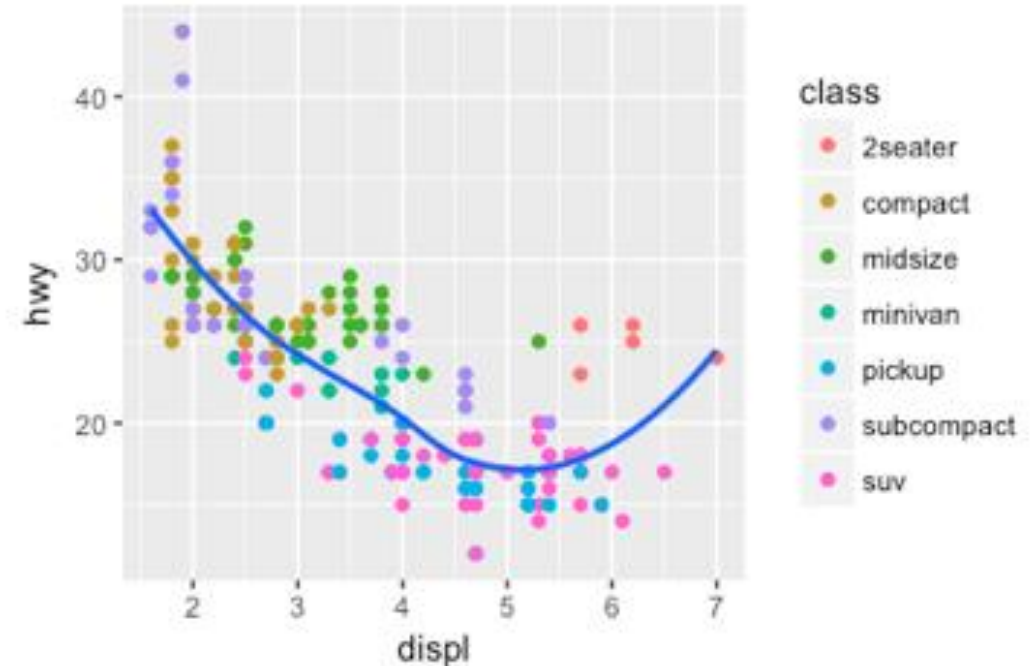


How does it work? Geometric shapes and aesthetic mappings

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 specific layer (i.e. `geom_point`).



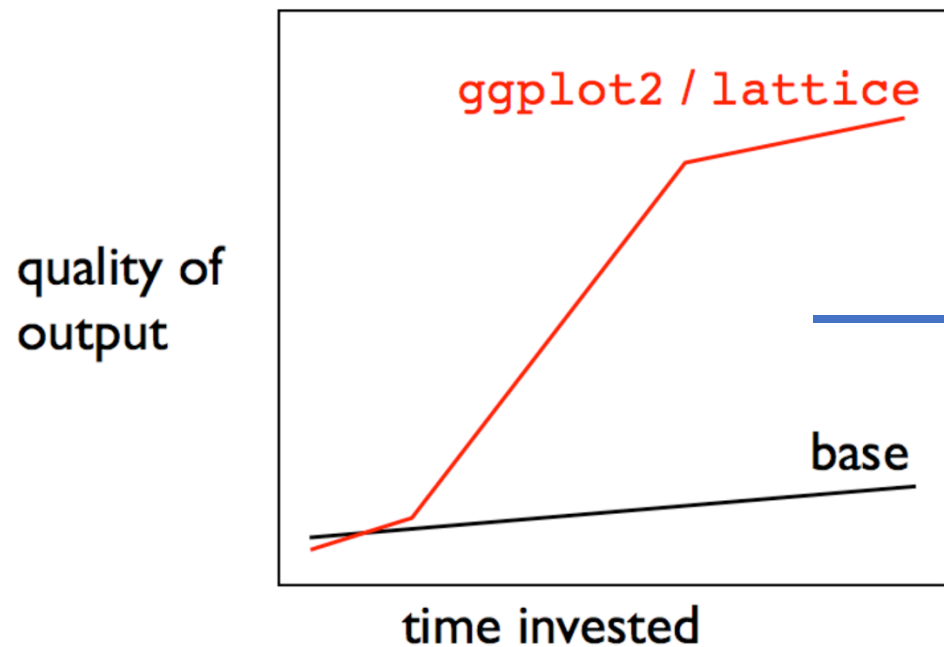
```
# color aesthetic passed to each geom layer
ggplot(mpg, aes(x = displ, y = hwy, color = class)) +
  geom_point() +
  geom_smooth(se = FALSE)
```



```
# 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)
```

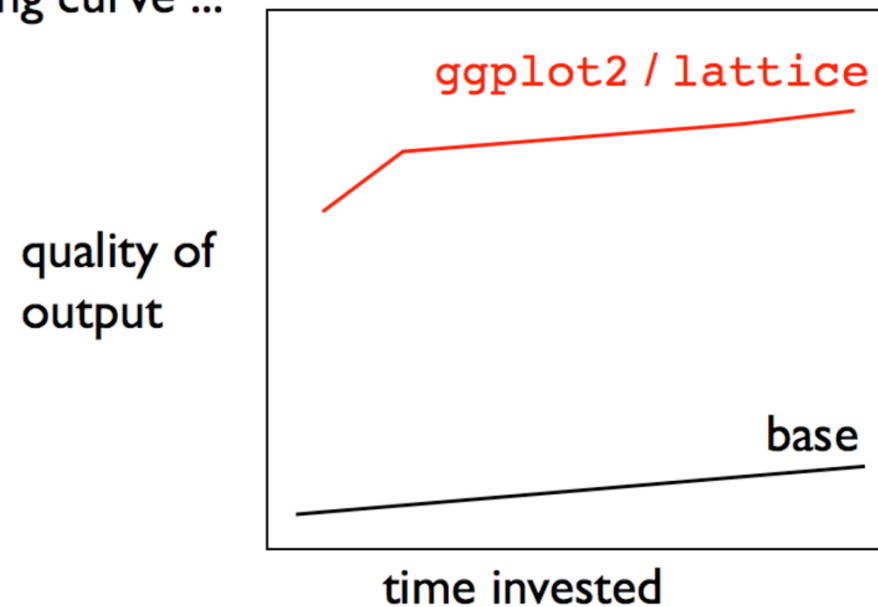
Why use ggplot2?

week one



* figure is totally fabricated but, I claim, it's true

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



Scales scale_.

Statistical transformations stat_:

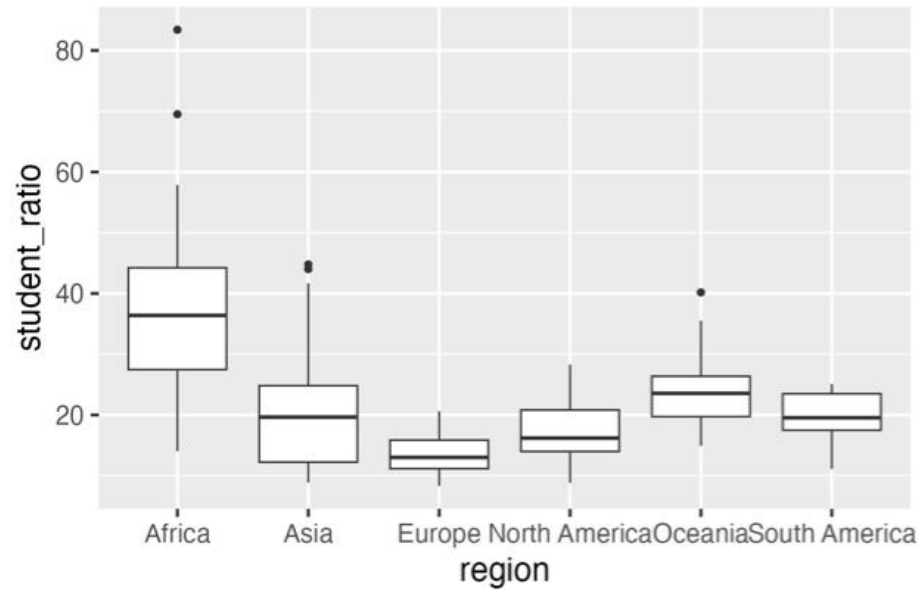
Coordinate system coord_:

Facets facet_:

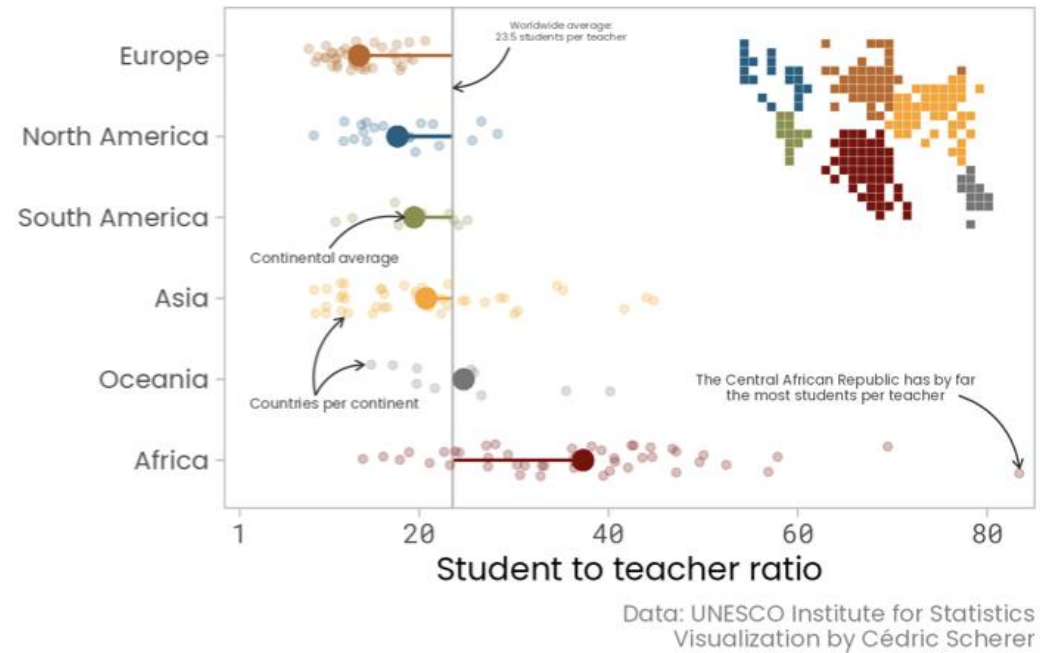
Visual themes theme():

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

Before

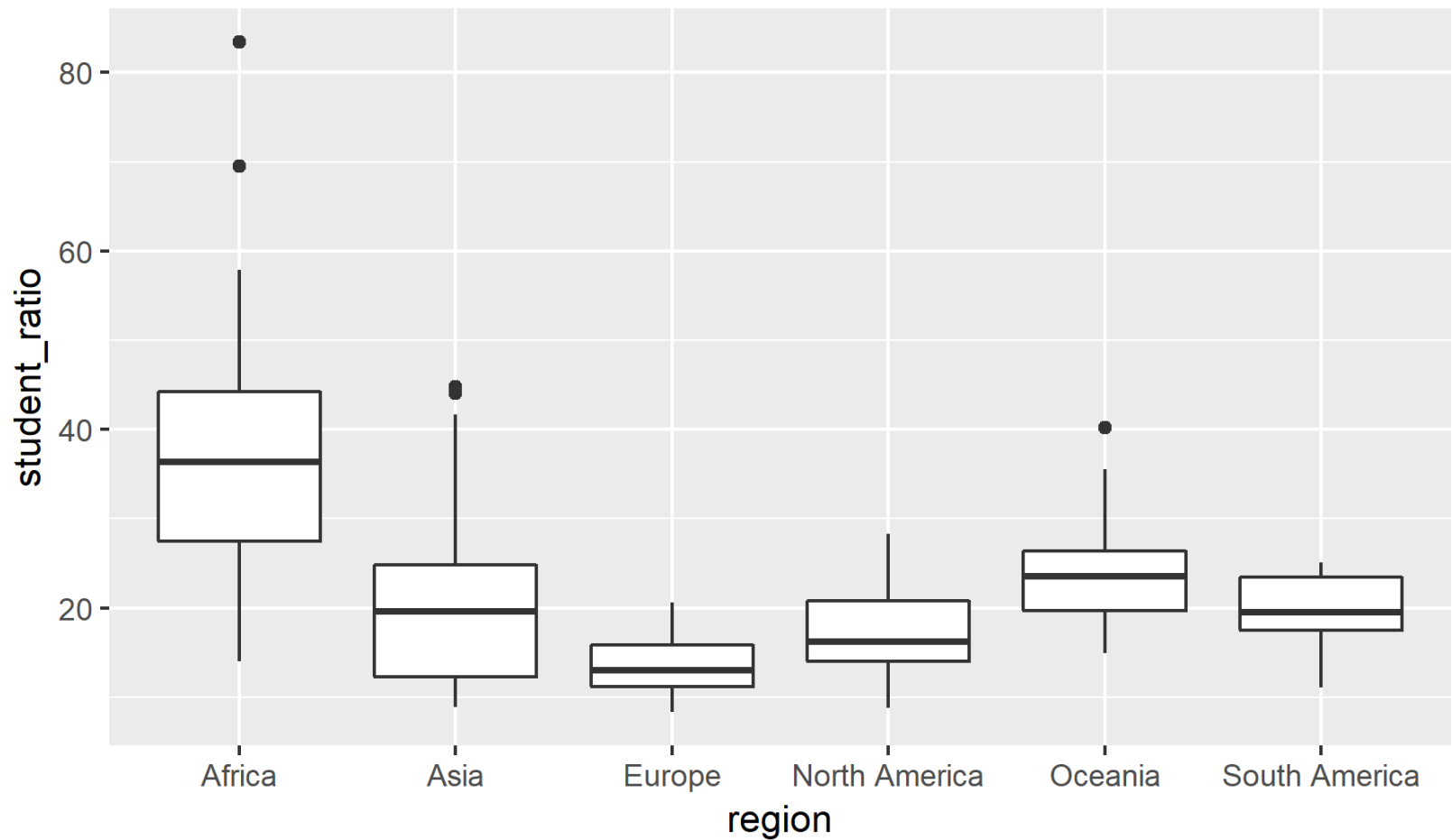


After



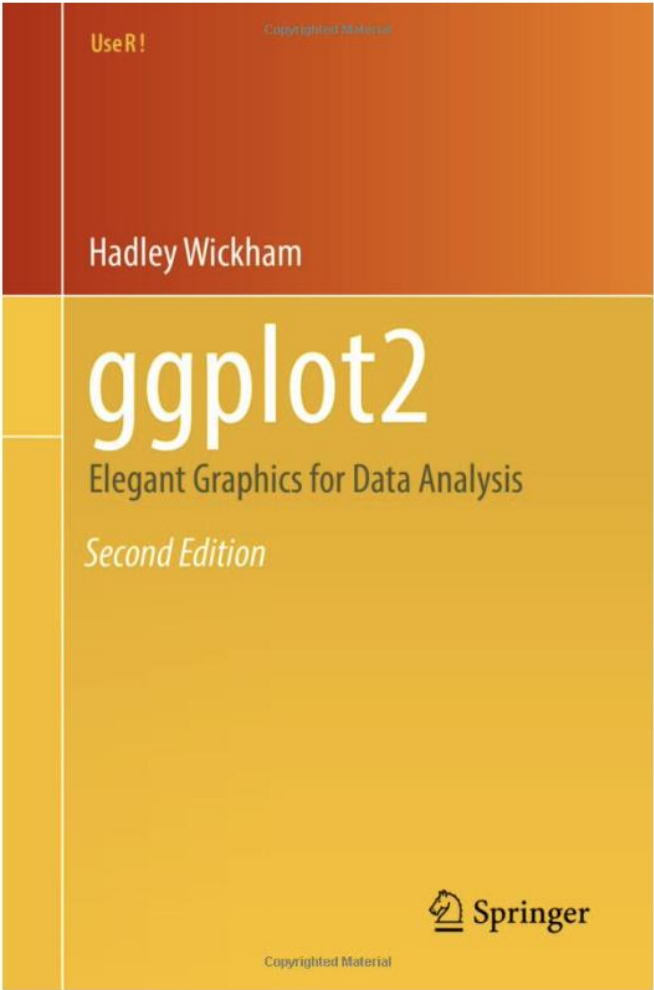
How does it work? Simple and powerful

The Evolution of a ggplot



Data: UNESCO Institute for Statistics
Visualization by Cédric Scherer

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](#)



Thomas Lin Pedersen
@ThomasLinPedersen · 5.77K subscribers · 2 videos
[More about this channel](#) >
[Subscribe](#)

Home Live Playlists Community

Uploads ▶ Play all



ggplot2 workshop part 2
78K views · Streamed 3 years ago



ggplot2 workshop part 1
205K views · Streamed 3 years ago

[Youtube ggplot2 tutorial](#)

ggplot2 extensions - gallery

[Add Your Extension!](#) [exts.ggplot2.tidyverse.org](#)

132 registered extensions available to explore

Sort: Github stars

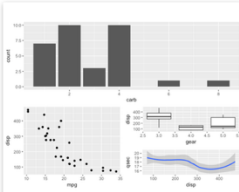
Text Filter: search name, author, descri

Author Filter:

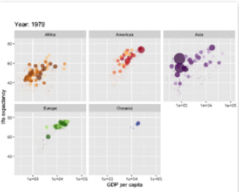
Tag Filter:

CRAN Only

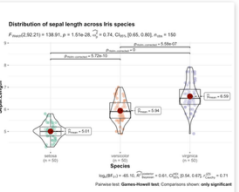
Showing 110 of 132



patchwork ⭐ 2347
Easy composition of ggplot plots using arithmetic operators
• author: thomasp85
• tags: visualization, composition
• js libraries:



gganimate ⭐ 1913
A Grammar of Animated Graphics.
• author: thomasp85
• tags: visualization, general
• js libraries:



ggstatsplot ⭐ 1669
'ggstatsplot' provides a collection of functions to enhance 'ggplot2' plots with results from statistical tests.
• author: IndrajeetPatil
• tags: visualization, statistics
• js libraries:



esquisse ⭐ 1712
Explore and Visualize Your Data Interactively with ggplot2
• author: dreamrs
• tags: visualization, interface
• js libraries:

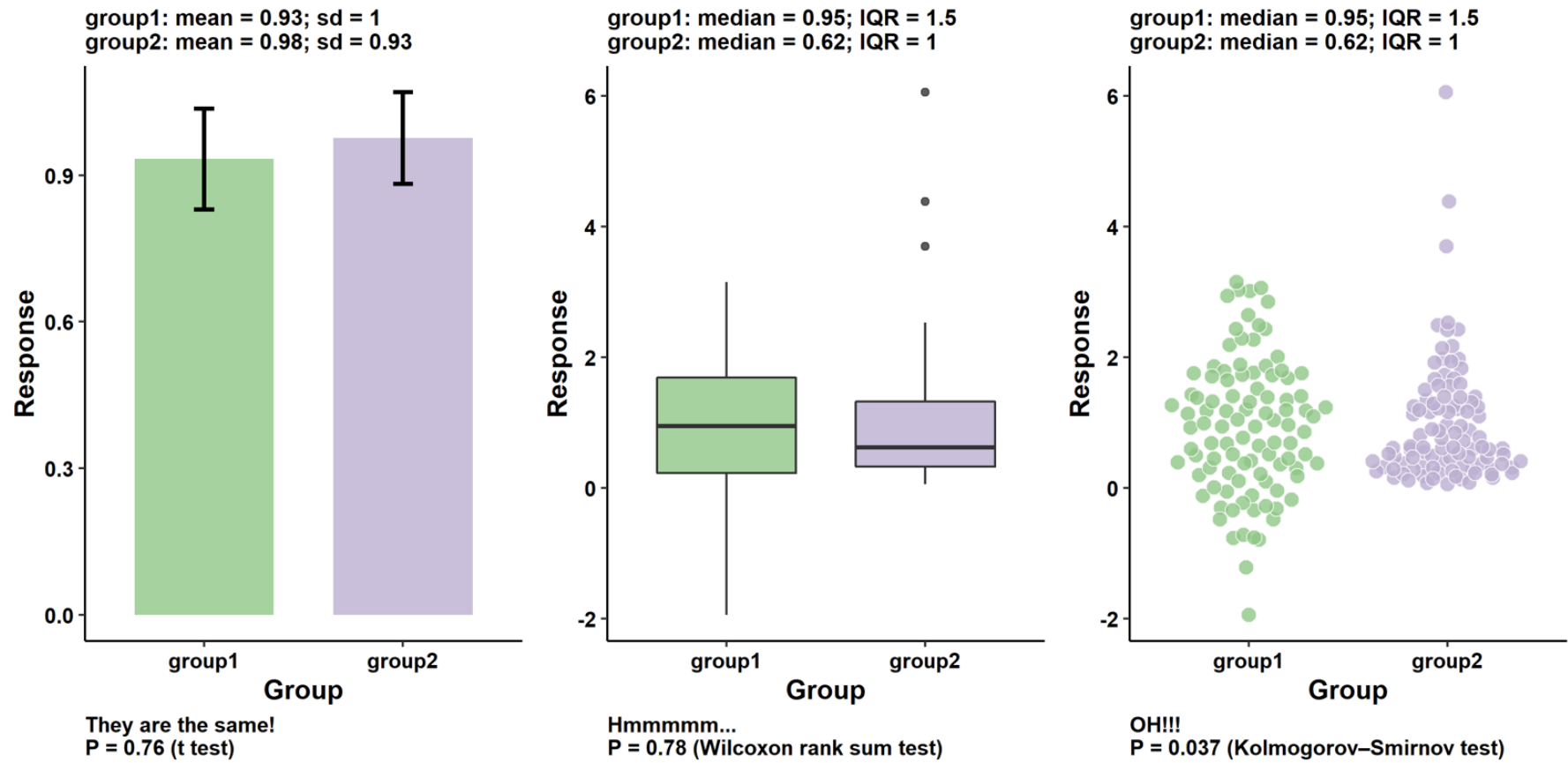
22

Awesome resources

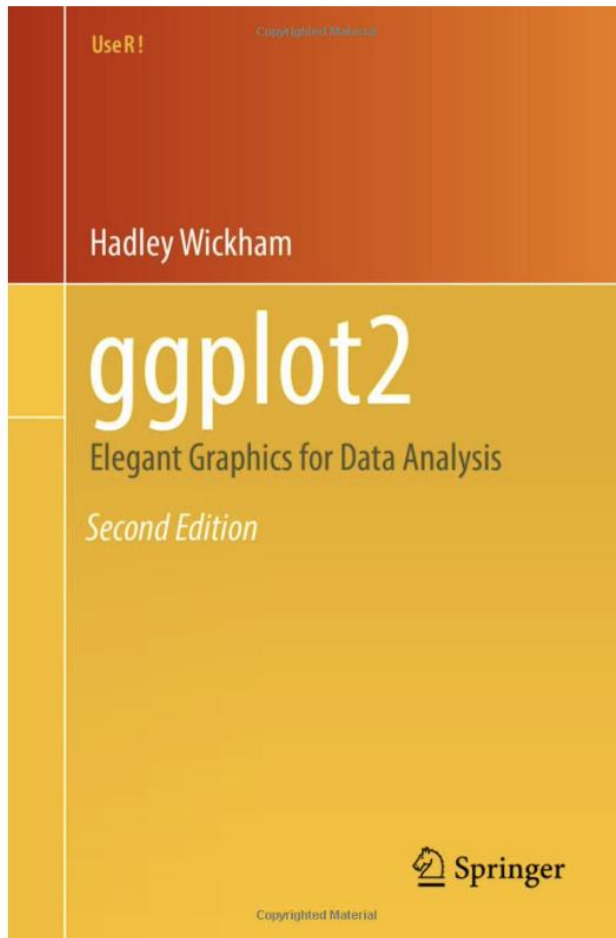
- <https://github.com/erikgahner/awesome-ggplot2>
- <https://r-graph-gallery.com/index.html>
- <http://www.sthda.com/english/wiki/ggplot2-essentials>
- Very good tricks: https://github.com/teunbrand/ggplot_tricks, https://albert-rapp.de/posts/ggplot2-tips/21_common_ggplot_mistakes/21_common_ggplot_mistakes.html
- <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



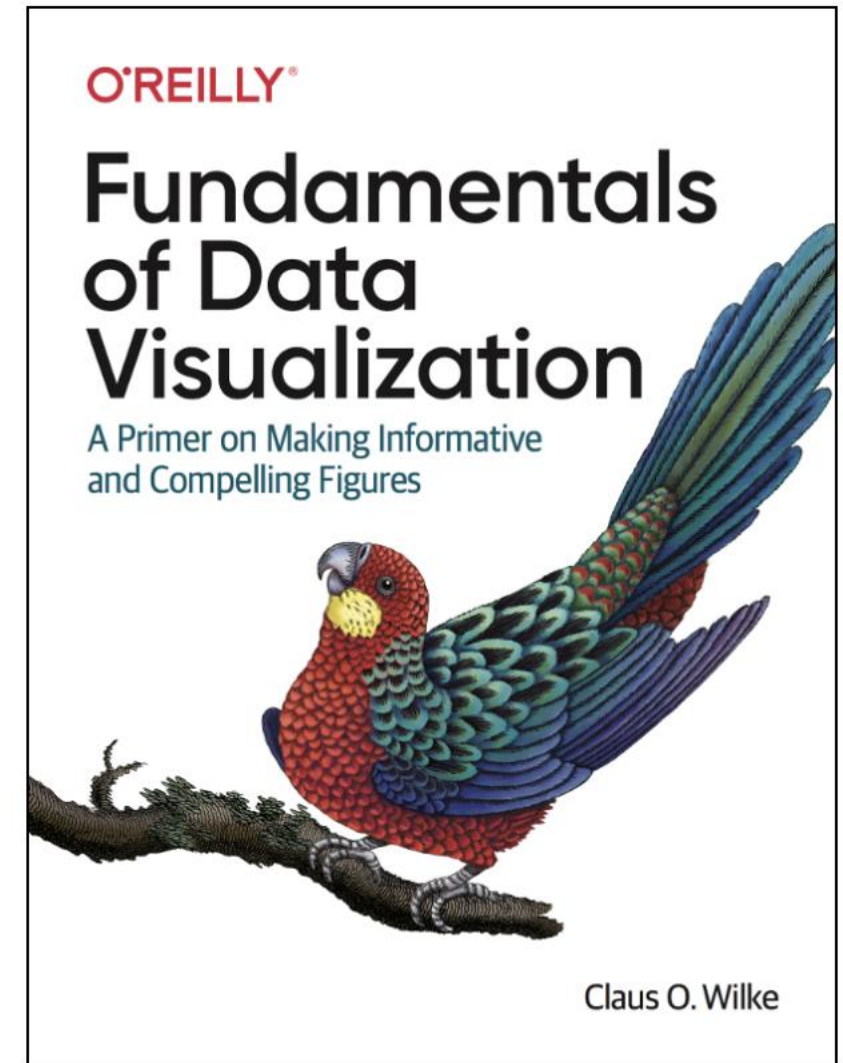
Learn more! Power in community



Core developers of ggplot2 are:

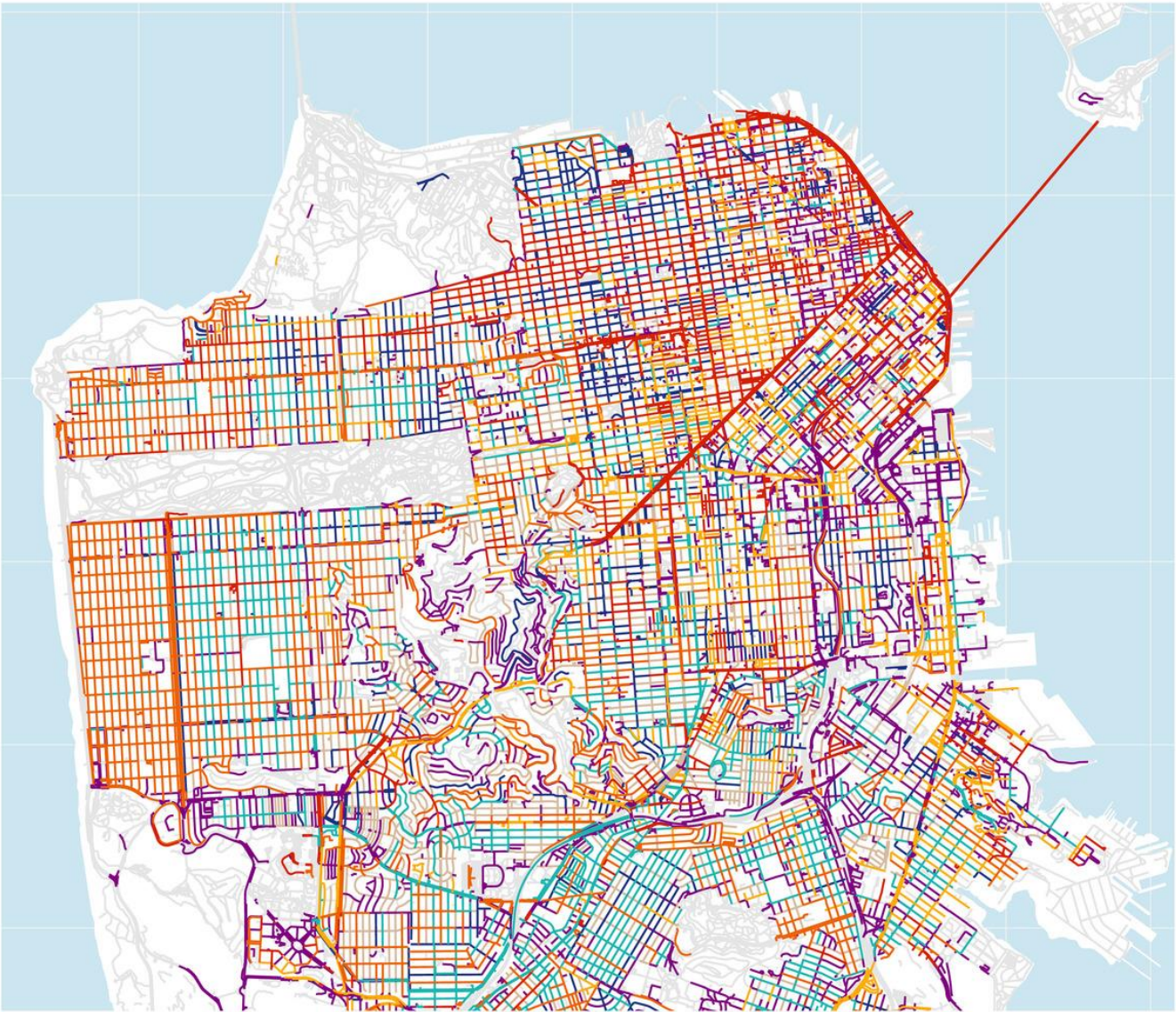
- [Winston Chang](#)
- [Lionel Henry](#)
- [Thomas Lin Pedersen](#)
- [Kohske Takahashi](#)
- [Claus Wilke](#)
- [Kara Woo](#)
- [Hiroaki Yutani](#)
- [Dewey Dunnington](#)

<https://ggplot2-book.org>



<https://clauswilke.com/dataviz/>

THE DOMINANT TREE SPECIES' PLANTED ALONG SAN FRANCISCO'S ROADS



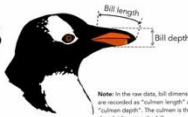
PLATANUS X HISPANICA (SYCAMORE LONDON PLANE)
PETROSIDEROS EXCELSA (NEW ZEALAND CHRISTMAS TREE)
LOPHOSPERMUM CONFERTUS (CHRISANTHEMUM)
TRISTANDOPSIS LAURINA (SWAMP MYRTLE)
PITTOSPORUM UNDULATUM (VICTORIAN BOX)
PRUNUS CERASIFERA (CHERRY PLUM)
OTHER TREE SPECIES

VISUALIZATION BY CÉDRIC SCHERER • DATA: DATASF • MAP: DATASF & OPENSTREETMAP CONTRIBUTORS

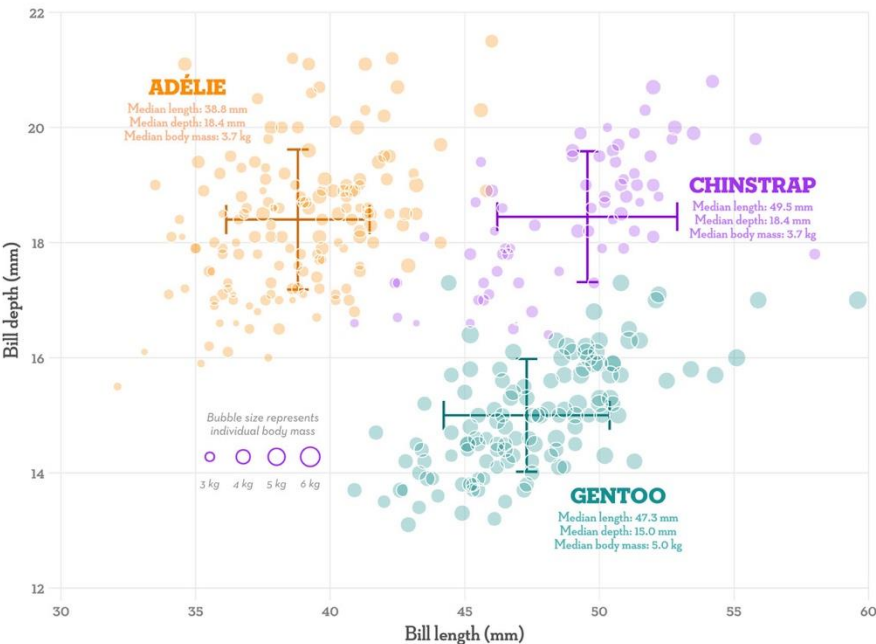
Stories!

BILL DIMENSIONS OF BRUSH-TAILED PENGUINS

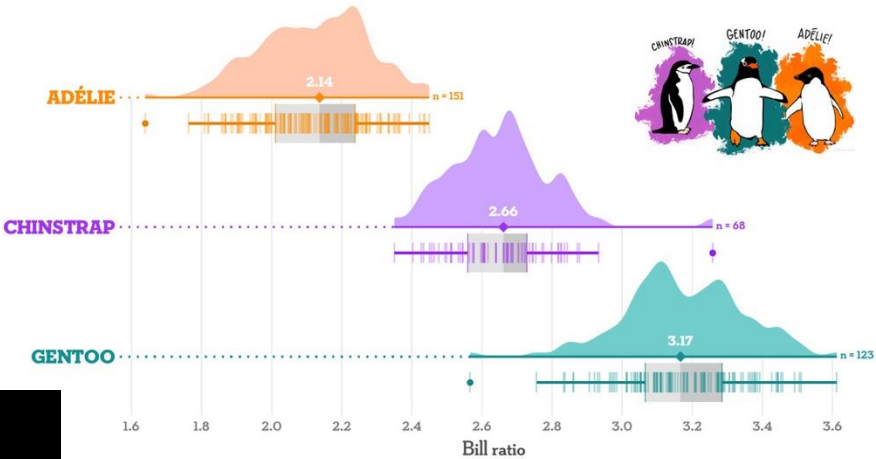
Pygoscelis adeliae (Adélie penguin) • *P. antarctica* (Chinstrap penguin) • *P. papua* (Gentoo penguin)



A. Scatterplot of bill length versus bill depth (median +/- sd)



B. Distribution of the bill ratio, estimated as bill length divided by bill depth



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

TidyTuesday

A weekly data project in R from the
R4DS online learning community

variables observations values

TidyTuesday

A weekly data project in R from the
R4DS online learning community

variables observations values

TidyTuesday

A weekly data project in R from the
R4DS online learning community

variables observations values

TidyTuesday

A weekly data project in R from the
R4DS online learning community

variables observations values

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:

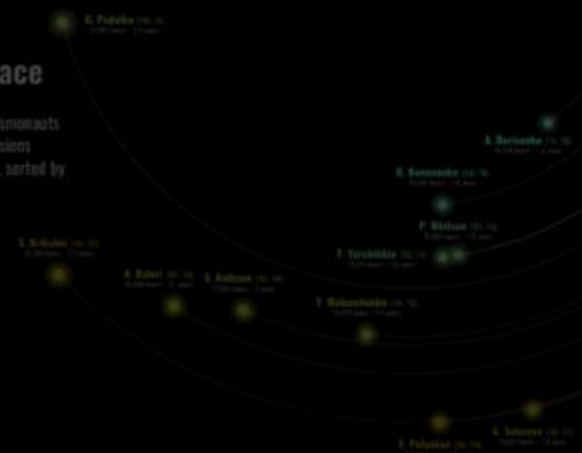
Email Majordomo@lists.bham.ac.uk with **subscribe uobrug** in the main body of a plain text email (please include a new line at the end and no other content, e.g. signatures).

Data VISUALIZATION using GGPLOT2

Laura Bravo

Travelling to Outer Space

Cumulative time in outer space for all 565 cosmonauts and astronauts who participated in space missions between April 23, 1961 and January 15, 2020, sorted by the year of their first mission.

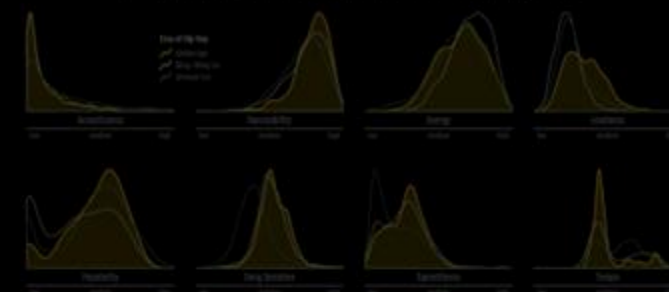


Visualization by Cedric Scherer - Data by Stawickich & Corbett 2020 (DOI: 10.17632/85tmmv2w1)

The cosmonauts and astronauts are sorted by the year of their first mission, and are placed equidistant. Thus, the width of a decade depends on the number of cosmonauts and astronauts during that period.

The Golden Age of Hip Hop in the Era of Spotify

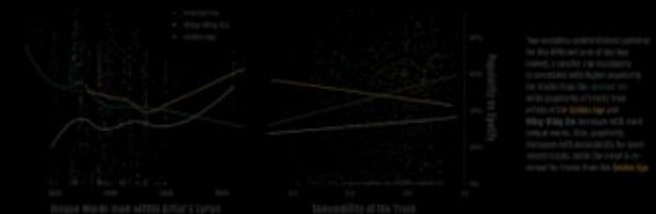
It is generally accepted that the **Golden Age of Hip Hop** occurred from the mid 1980s to the mid 1990s. It was then that all the elements of the culture—b-boying, graffiti art, fashion, and rap—broke cover to enter the mainstream. K.R.A., Eric B. & Rakim, Run DMC, and the Beastie Boys cleaned rap music to become the culture's crowning glory. With the likes of DMX, Dr. Dre, Jay-Z, Nas, and 50 Cent all selling albums to their tens of millions, Hip Hop became a genre changer, one of the most popular styles in modern music and reinvigorated youth culture.



Hip-hop during the **Golden Age (1985-1995)** was more an artist's time than a producer's. Producers often were themselves and better told of their stories and experiences. During 1995-2000, and the 2000s, producers became more specialized in the culture, focusing on specific styles. Today, though, the **Golden Age** has been largely replaced by a more commercial sound, with a focus on mainstream appeal. More artists and labels have found their way to the top of the charts.



Popular rappers from nowadays, however, can say fewer unique words. Is that the reason for the higher popularity?



The Spotify Effect on Track Length?

Track lengths dropped from an average of 4:45 minutes in the 1970s to 3:45 minutes in the 2010s. Is this due to the Spotify effect? The effect can be seen in the fact that the average track length in the 2010s is 3:45 minutes, which is the same as the average track length in the 1970s.

