

Thomas Hütter

Data Innovation
Summit 2021

50 ways to show your data

„Condensed edition“



50 ways to show your data

Thomas Hütter, Diplom-Betriebswirt

- Application developer, consultant, accidental DBA, author
- Worked at consultancies, ISVs, end user companies
- SQL Server > 6.5, former „Navision“ > 3.0, R > 3.1.2
- Speaker at Data&Dev events around Europe

 @DerFredo <https://twitter.com/DerFredo>

 de.linkedin.com/in/derfredo

 www.xing.com/profile/Thomas_Huetter



SWE CON



sqlbits

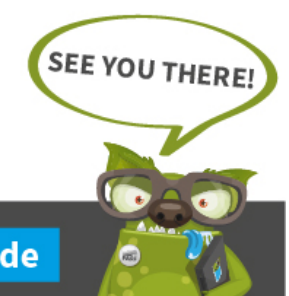


JOIN !
THE DATABASE CONFERENCE

SQLdays
konferenz

Pure expertise at the
SQL Server Konferenz 2018
FEB 26th - 28th, 2018 | DARMSTADT

Sign up now at sqlkonferenz.de



Agenda



Agenda

- What's the fuzz about this R language?



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:
 - from facets



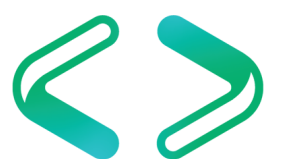
Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:
 - from facets
 - via waterfalls and radar



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:
 - from facets
 - via waterfalls and radar
 - to animated plots and yet some extras



Agenda

- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:
 - from facets
 - via waterfalls and radar
 - to animated plots and yet some extras
- Round-up, resources



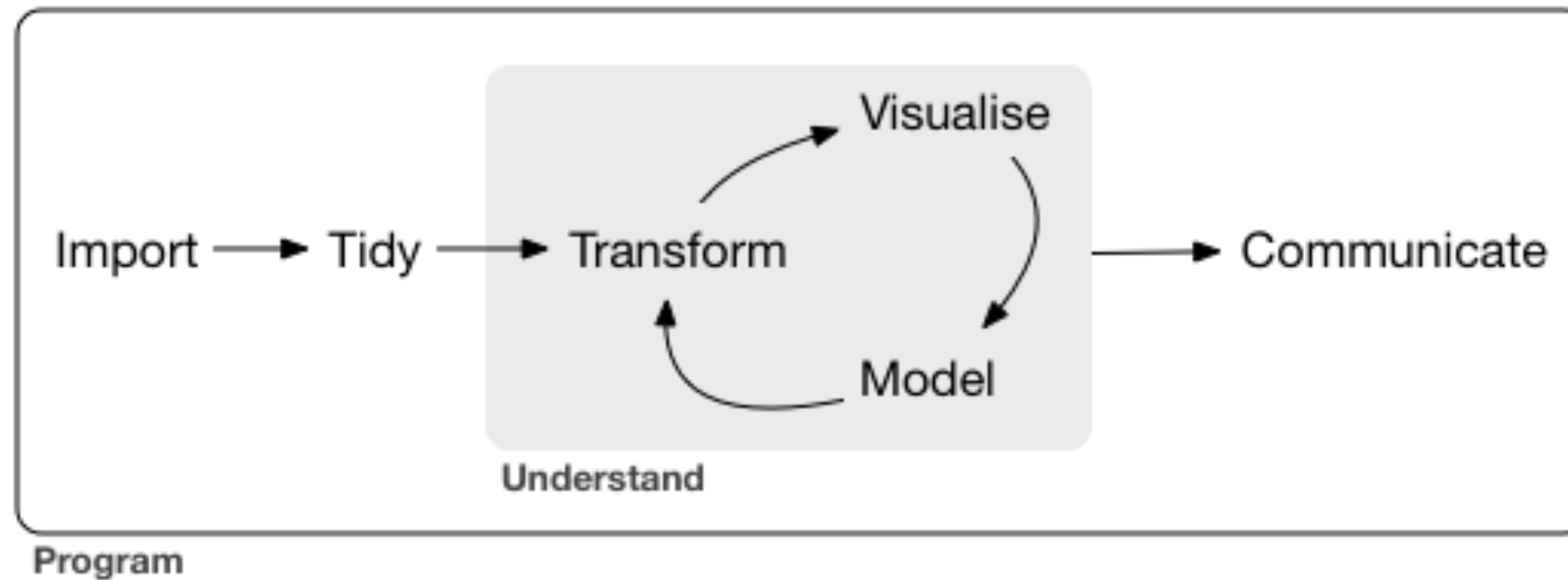
What's the fuzz about this **R** language?

- Programming language for statistical computing, analysis, visualization - widely used by statisticians, data miners, analysts, data scientists
- Created by **R**oss Ihaka and **R**obert Gentleman, Uni Auckland, in 1993 as an open source implementation of the (1970s) S language
- GNU project, maintained by the R Foundation for Statistical Computing, compiled builds für Mac OS, Linux, Windows, supported by R Consortium
- Extensible through user-created packages, > 18000 available on CRAN
- Commercial support, e.g. since 2007 by Revolution Analytics, acquired by Microsoft in 2015, now provide Microsoft R Open, R Server
- Support for R now built into SQL Server, R Visuals in Power BI, Azure ML, Data science VM



The philosophy of tidy data

What a typical data analysis/data science project may look like



The components of the Tidyverse (ggplot2 being one part) cover these tasks and can help you to accomplish them in a concise manner.

figure © 2017 Wickham/Grolemund: „R for Data Science“

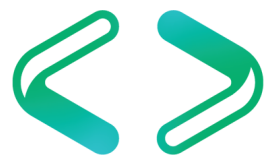
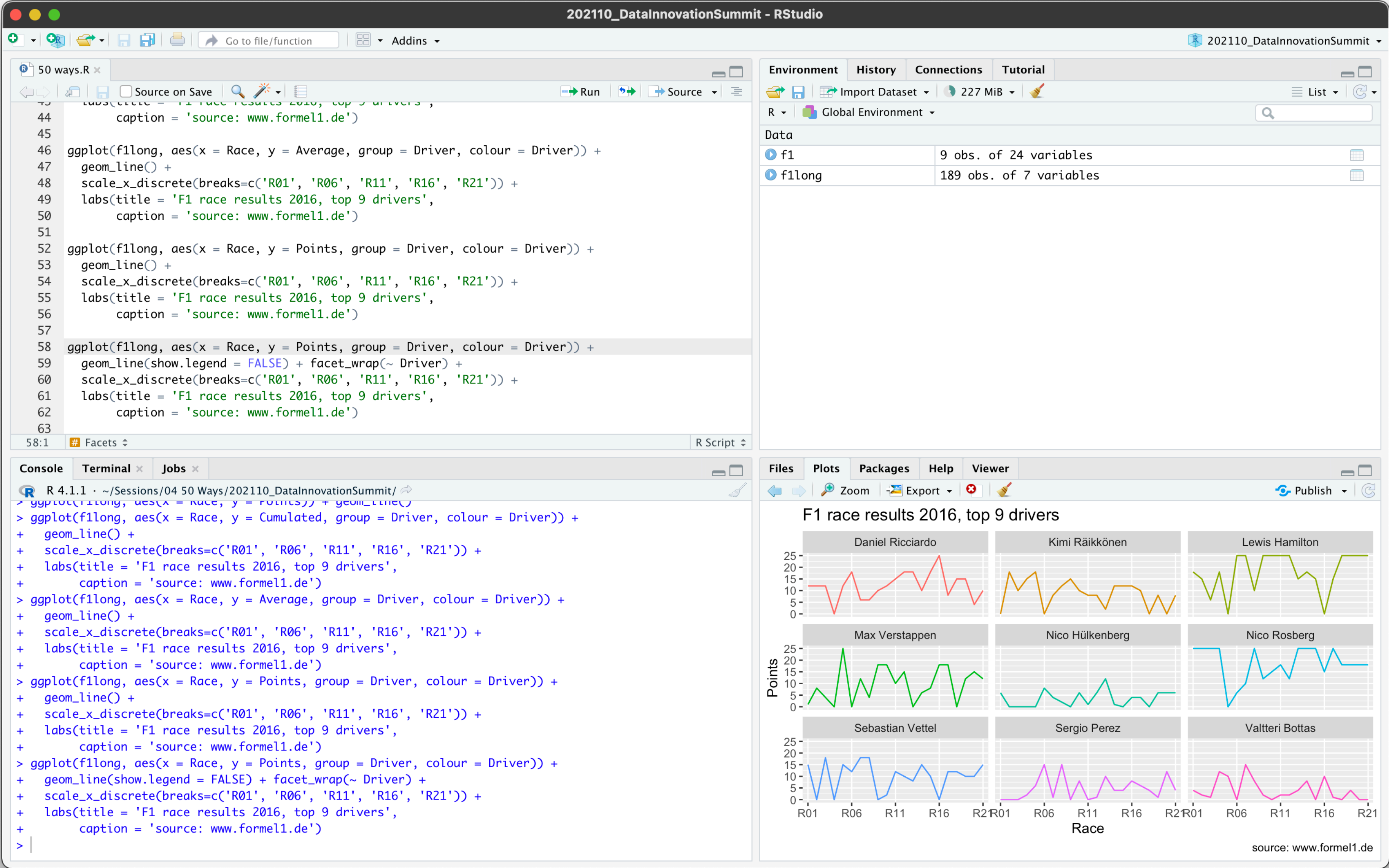


A grammar of graphics: ggplot2

- „The grammar of graphics“, a 2005 book by Leland Wilkinson et al. served as a foundation for implementing the R package ggplot2
- My simple approach to „what is needed to describe a graph“:
 - data: *what do you want to show*
 - aesthetic mappings: *which relations are there to show*
 - geometries: *how do you want to show it*
 - extras:
 - guides: axes, scales and legends
 - labels, annotations
 - facetting, coordinate systems
 - colours, themes
 - ...



Introducing RStudio, the de-facto standard IDE



50 ways to show your data

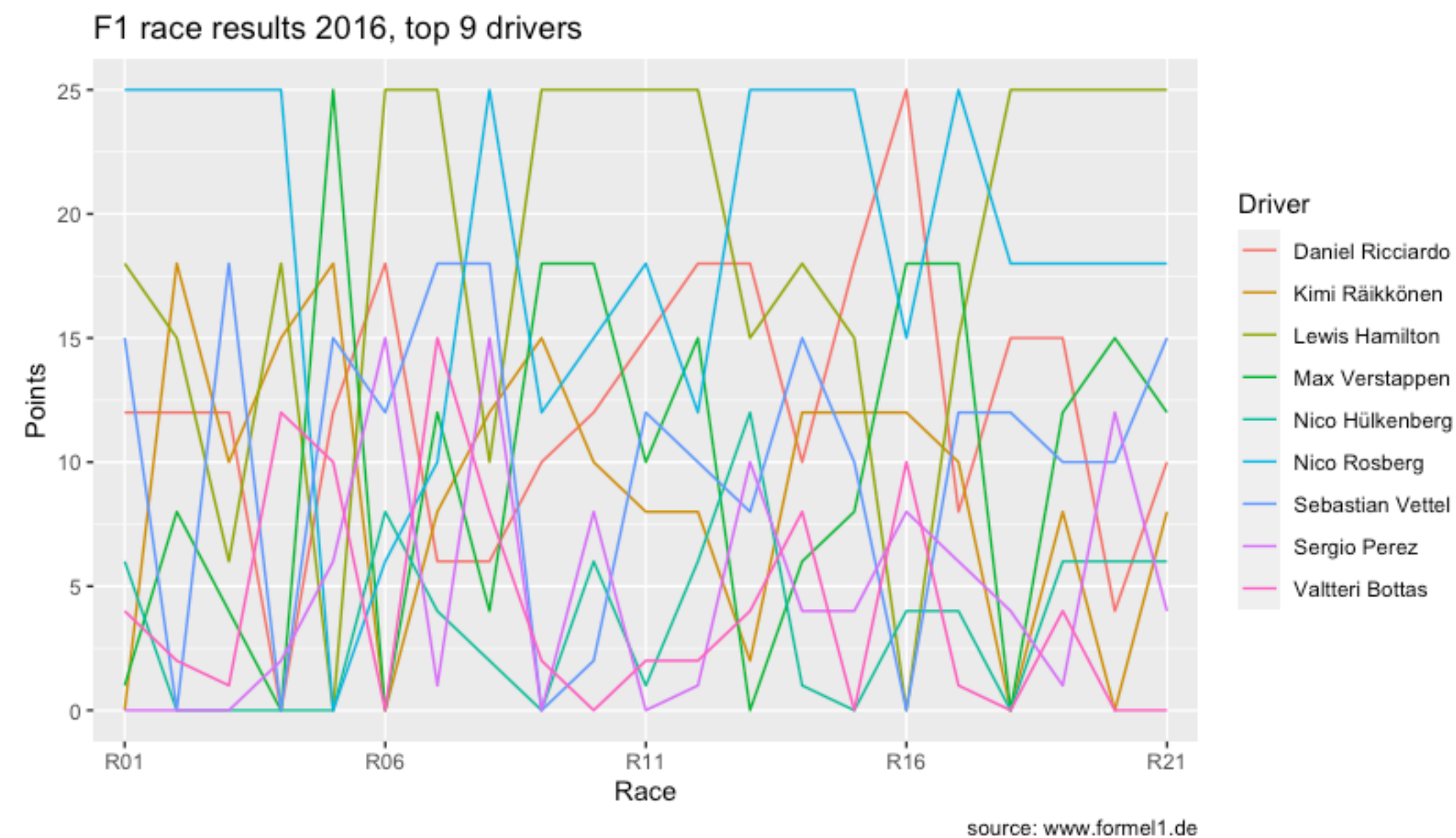
- facets
show small multiples,
apply the big picture to subsets of your data

```
f1 <- read_html('http://www.formel1.de/saison/wm-stand/2016/fahrer-wertung') %>%  
  html_node('table') %>%  
  html_table()  
  
colnames(f1) <- c('Pos', 'Driver', 'Total', sprintf('R%02d', 1:21))  
f1 <- as_tibble(f1) %>%  
  filter(as.integer(Pos) <= 9)  
f1$Driver <- as.factor(f1$Driver)  
f1[, -2] <- apply(f1[, -2], 2, function(x) as.integer(gsub('-', '0', as.character(x))))  
f1long <- gather(f1, Race, Points, R01:R21)
```



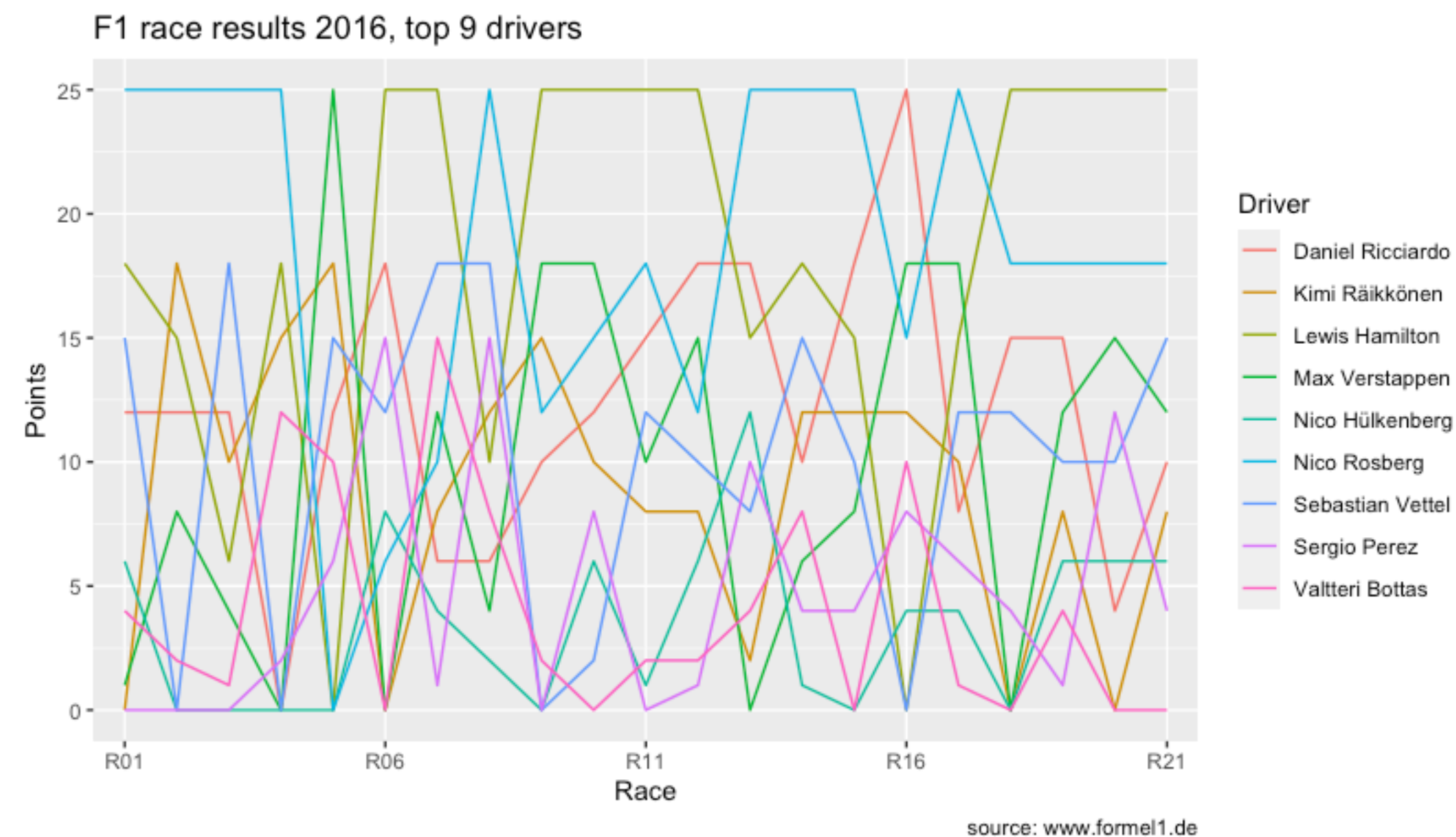
50 ways to show your data

- facets
show small multiples,
apply the big picture to subsets of your data



50 ways to show your data

- facets
show small multiples,
apply the big picture to subsets of your data

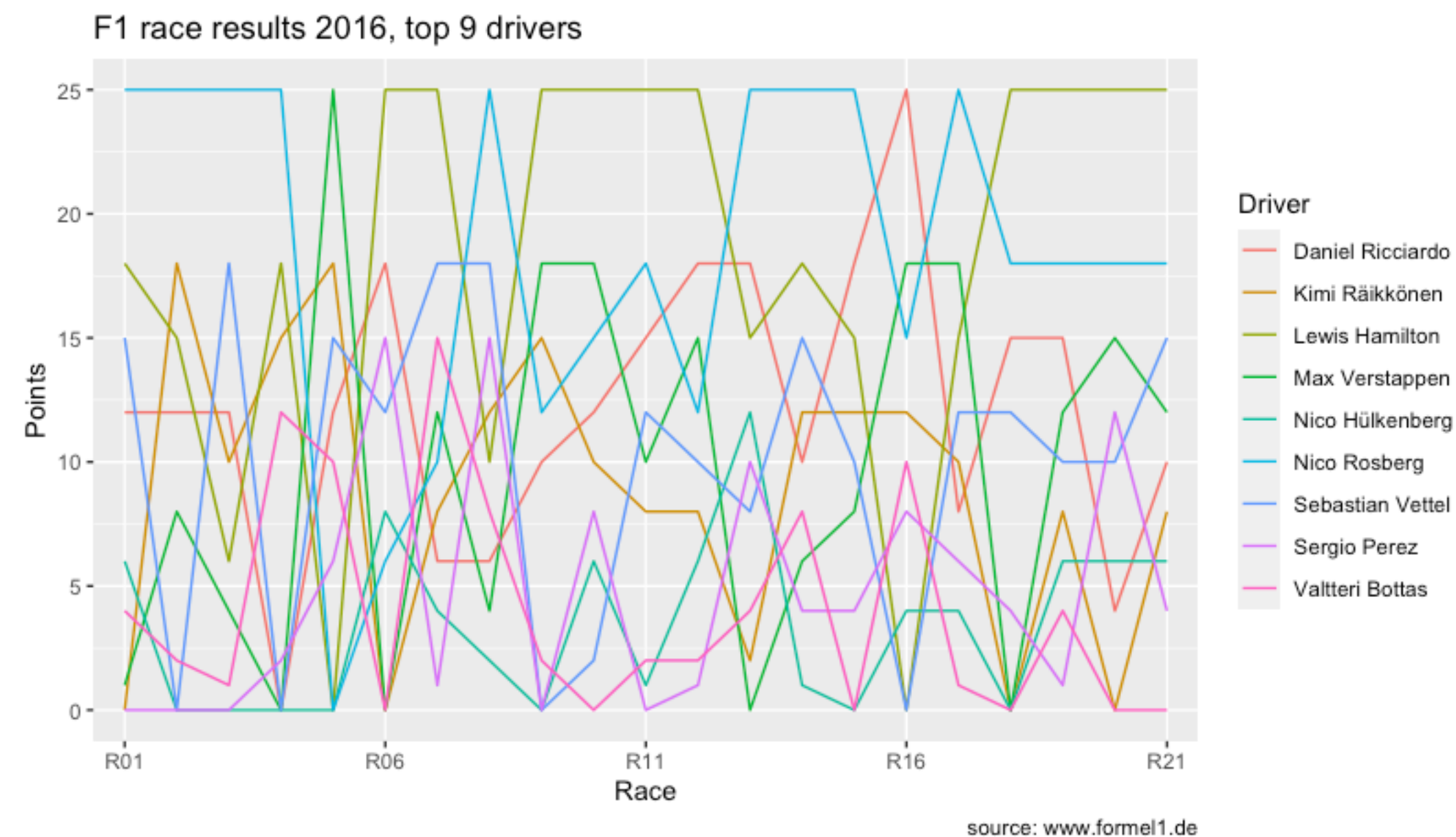


```
ggplot(f1long,  
  aes(x = Race, y = Points, group = Driver, colour = Driver)) +  
  geom_line() +  
  scale_x_discrete(breaks=c('R01', 'R06', 'R11', 'R16', 'R21')) +  
  labs(title = 'F1 race results 2016, top 9 drivers',  
    caption = 'source: www.formel1.de')
```

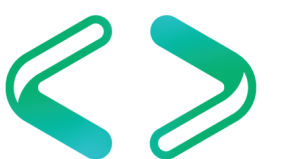


50 ways to show your data

- facets
show small multiples,
apply the big picture to subsets of your data

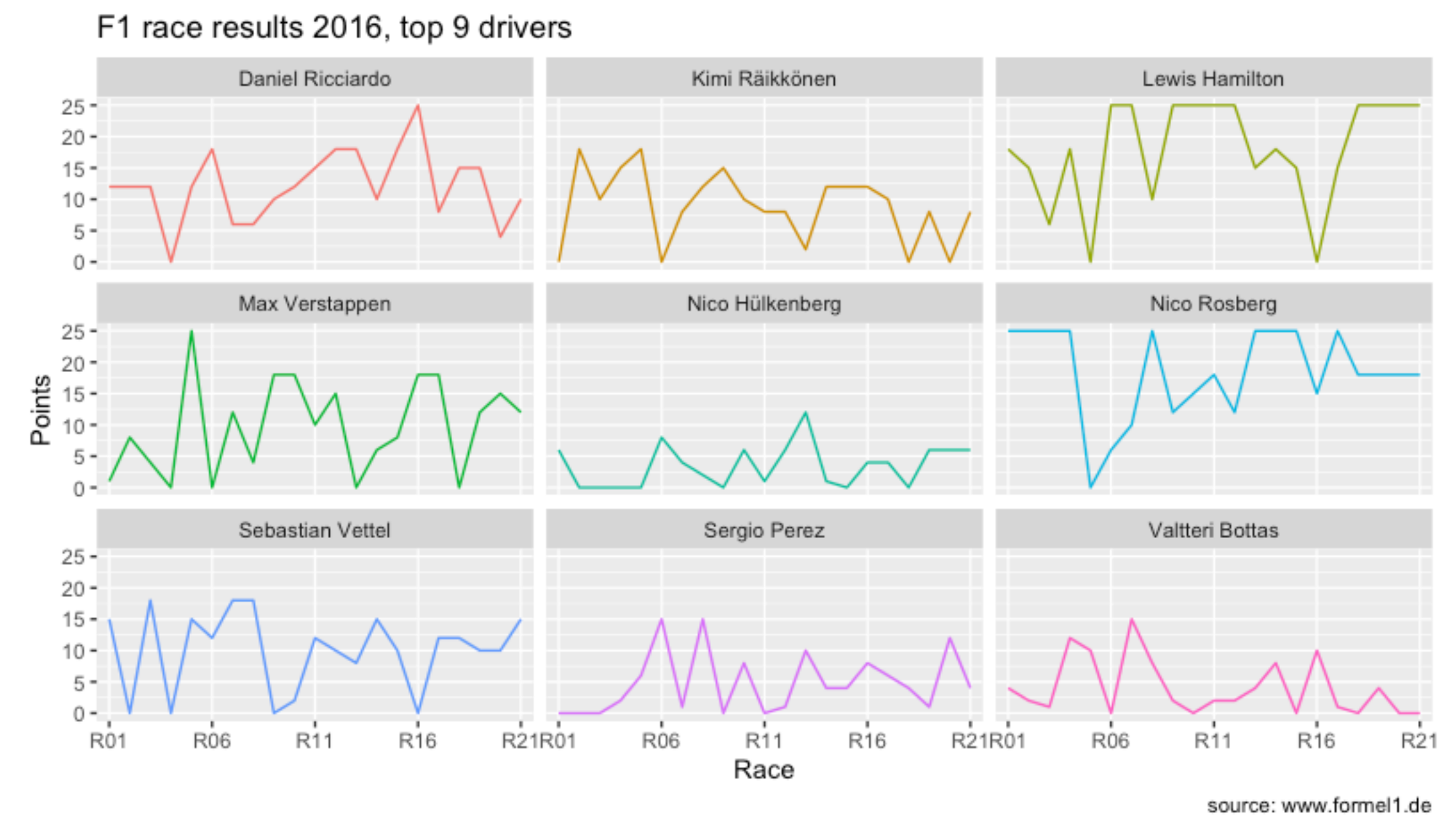
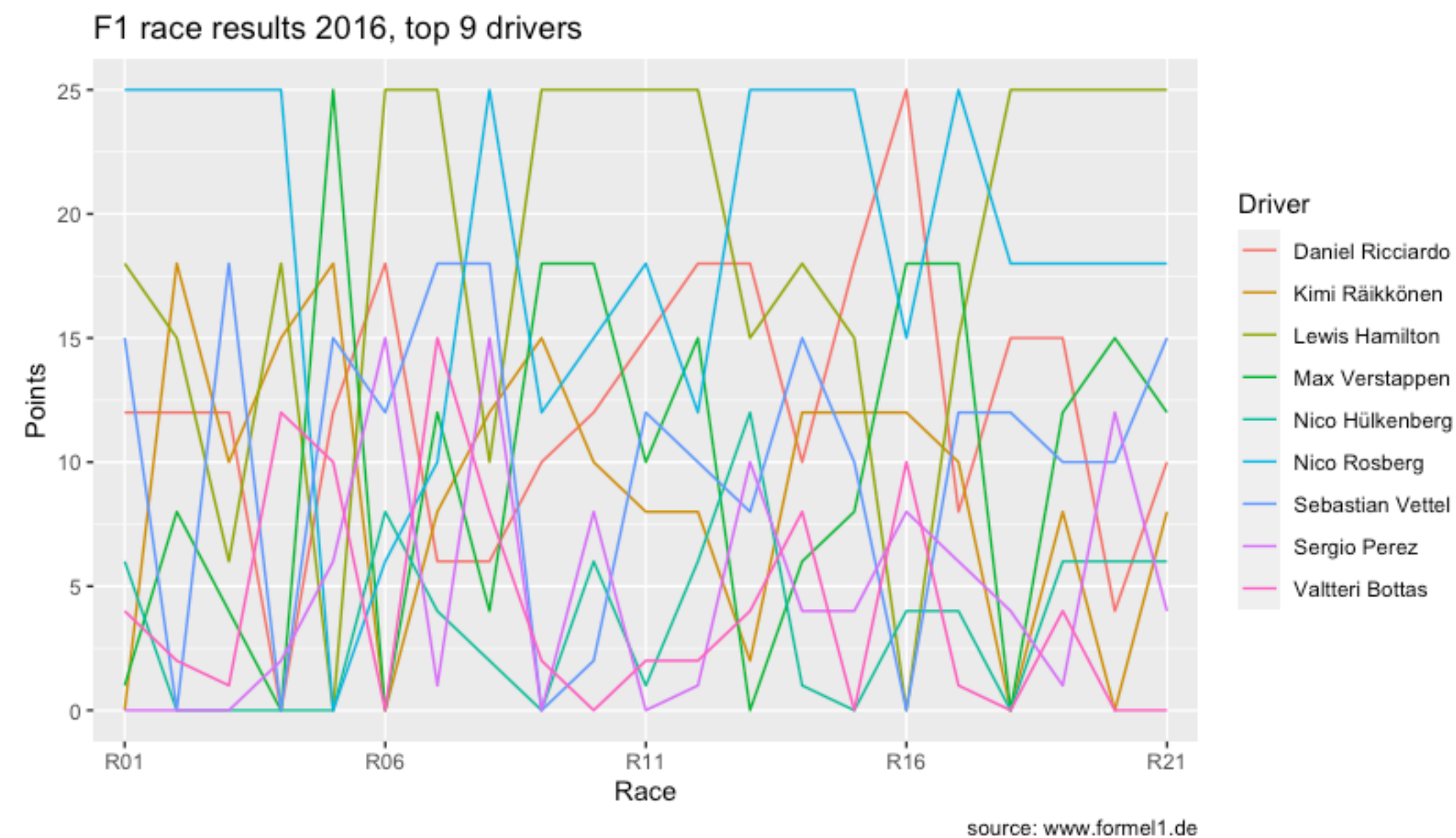


```
ggplot(f1long,  
  aes(x = Race, y = Points, group = Driver, colour = Driver)) +  
  geom_line(show.legend = FALSE) + facet_wrap(~ Driver) +  
  scale_x_discrete(breaks=c('R01', 'R06', 'R11', 'R16', 'R21')) +  
  labs(title = 'F1 race results 2016, top 9 drivers',  
    caption = 'source: www.formel1.de')
```



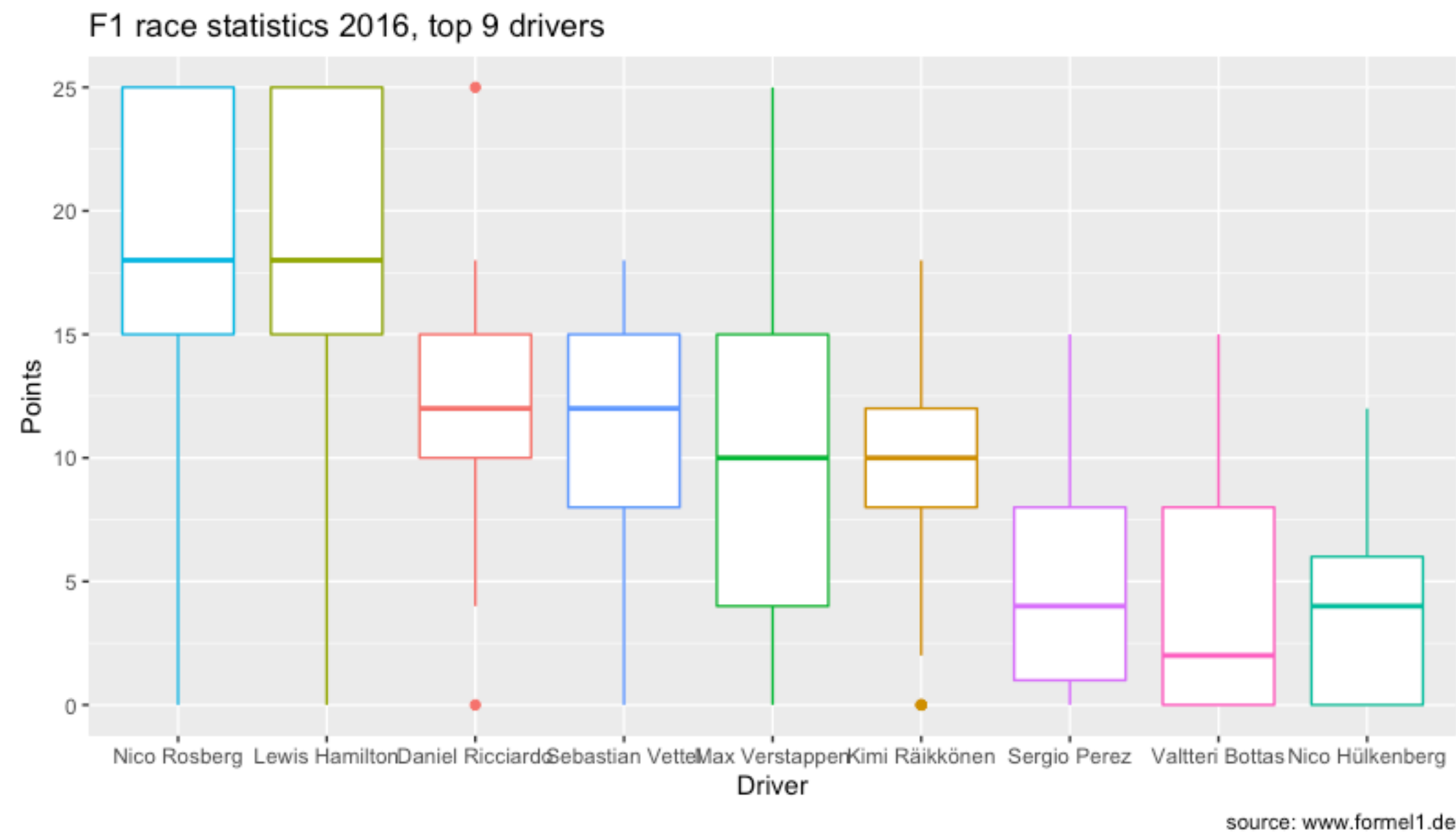
50 ways to show your data

- facets
show small multiples,
apply the big picture to subsets of your data



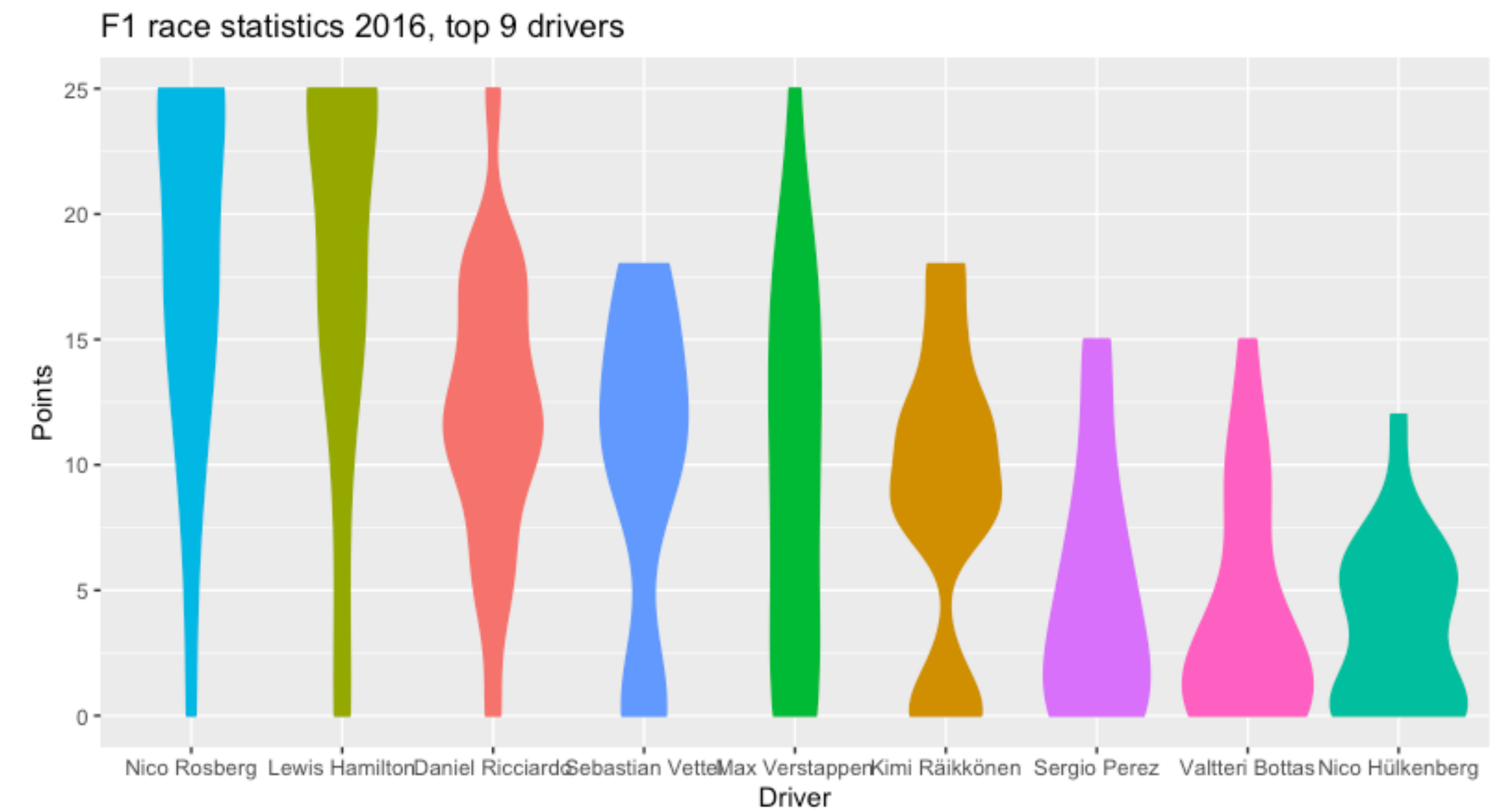
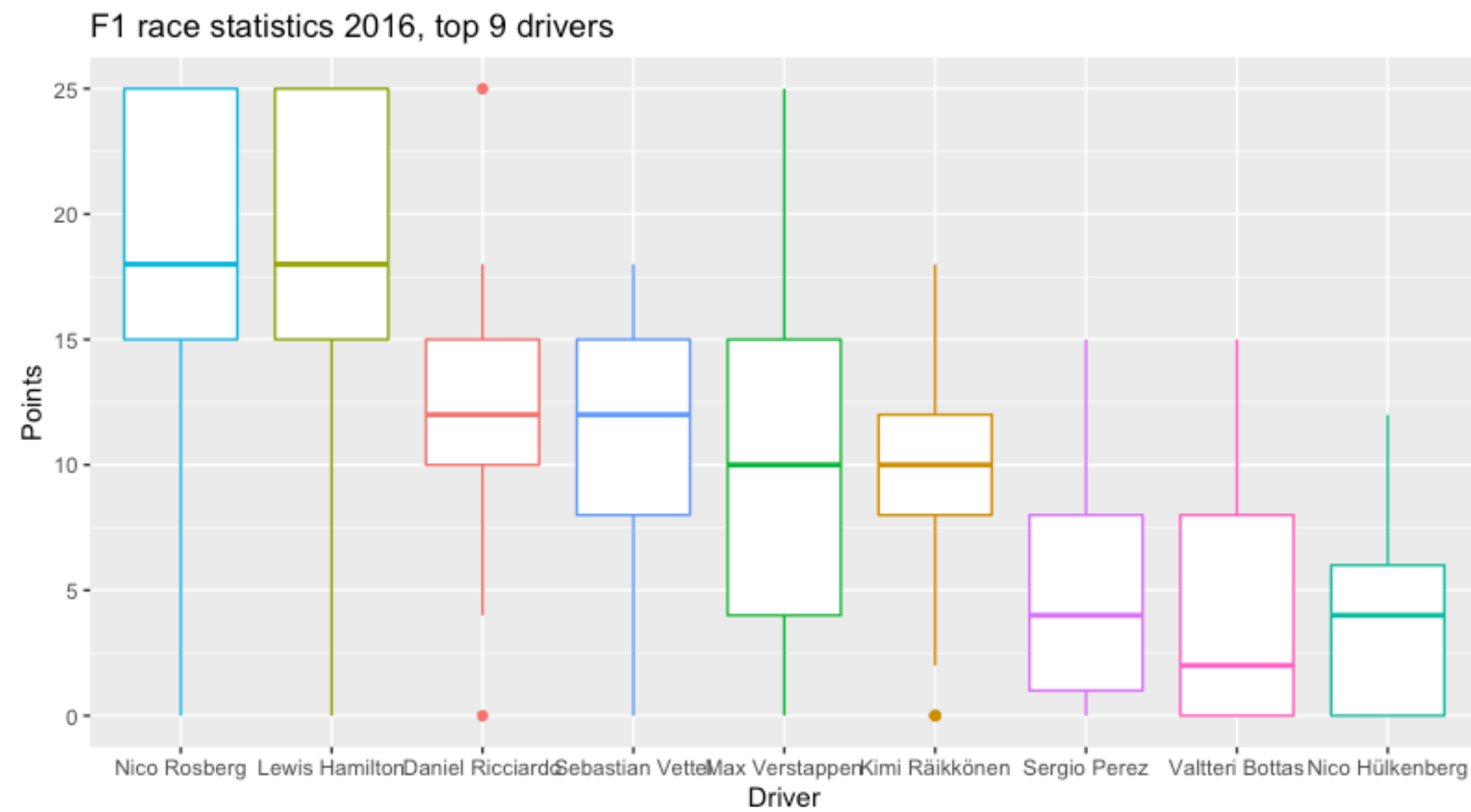
50 ways to show your data

- violins
show more detail of the data distribution than a box plot



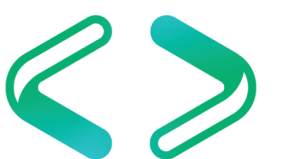
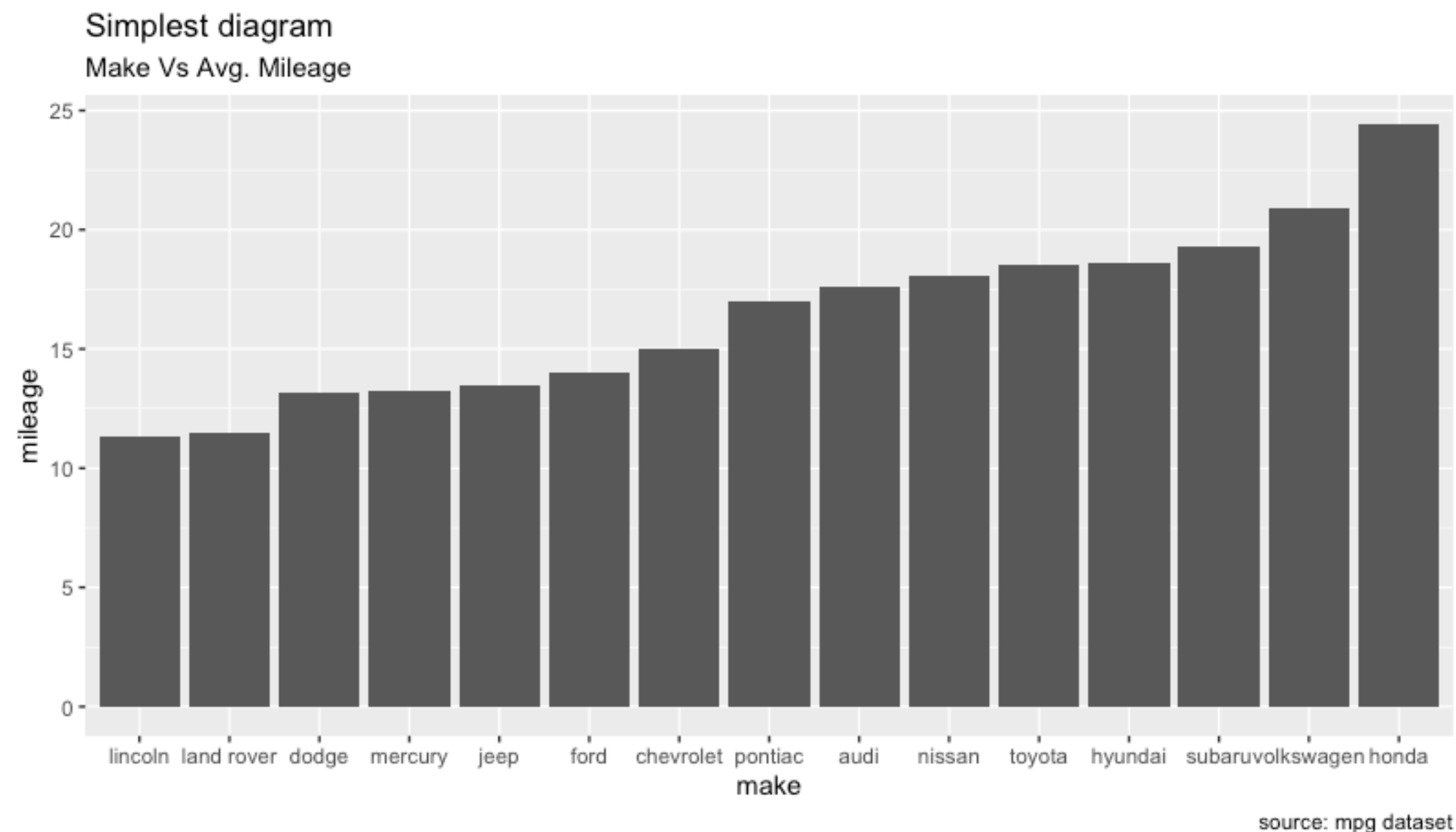
50 ways to show your data

- violins
show more detail of the data distribution than a box plot



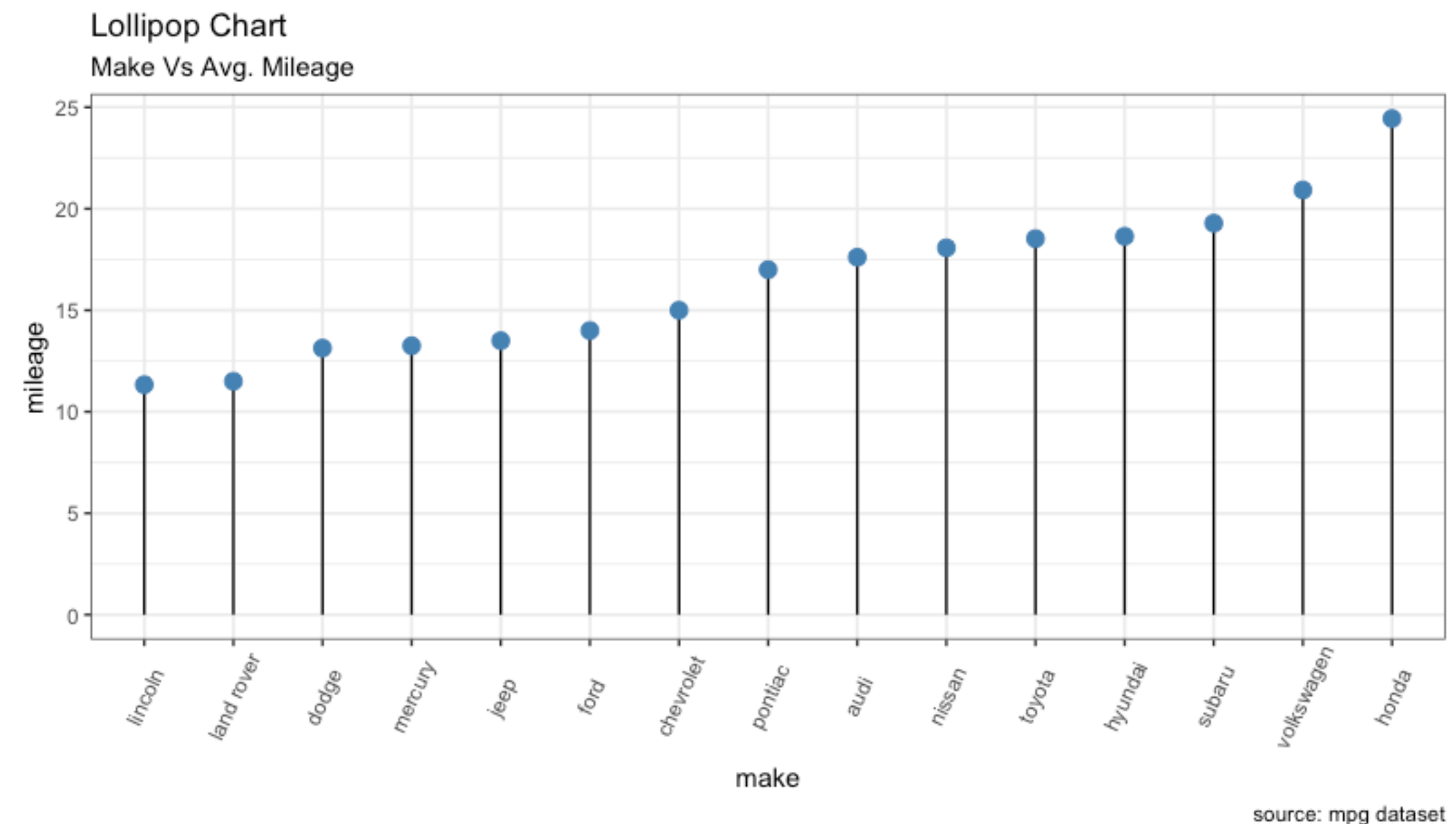
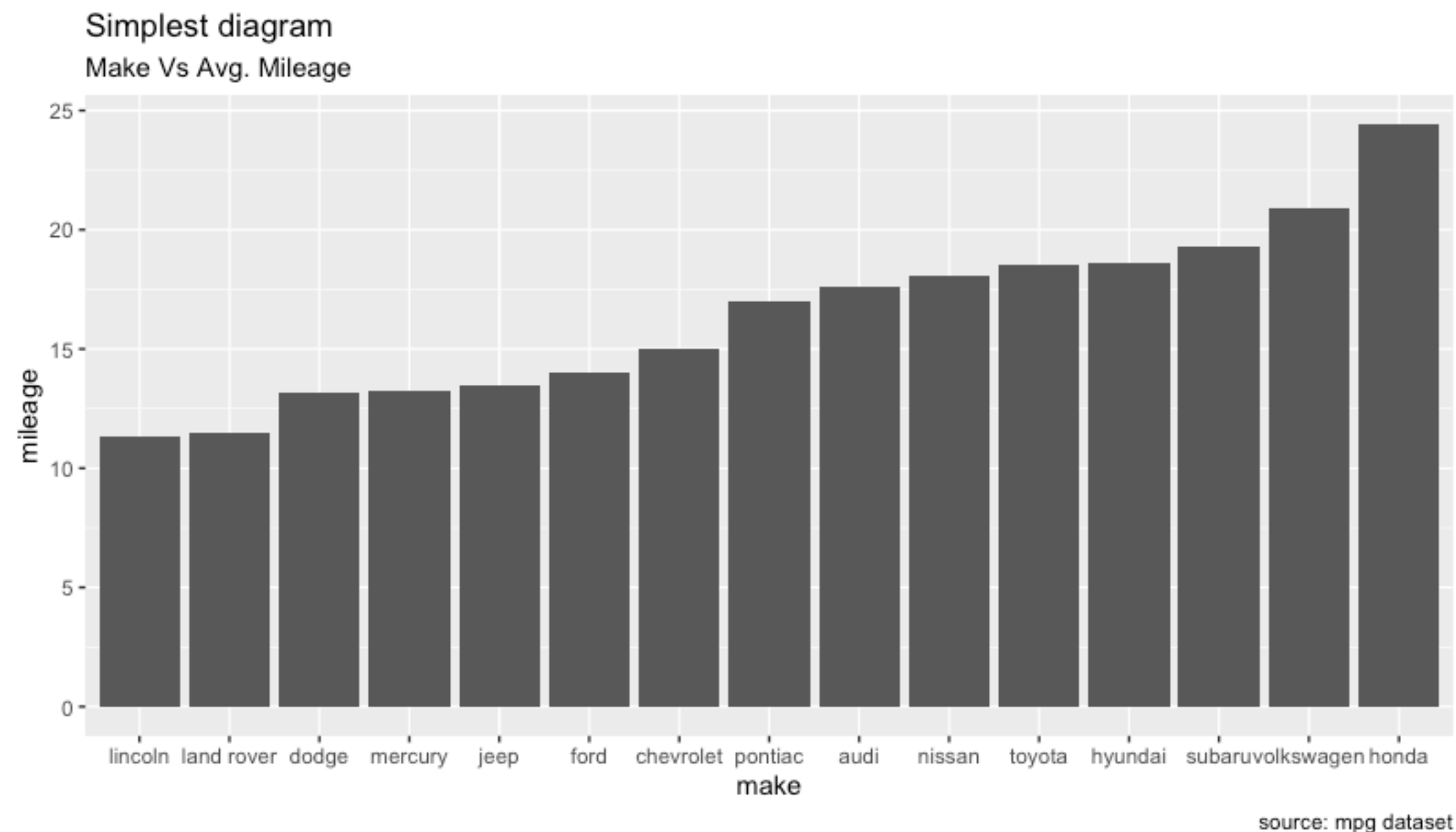
50 ways to show your data

- lollipop charts [ggalt]
alternative to charts with lots of nearly-similar-size bars,
on screen: reduce moiré, on paper: reduce waste of ink



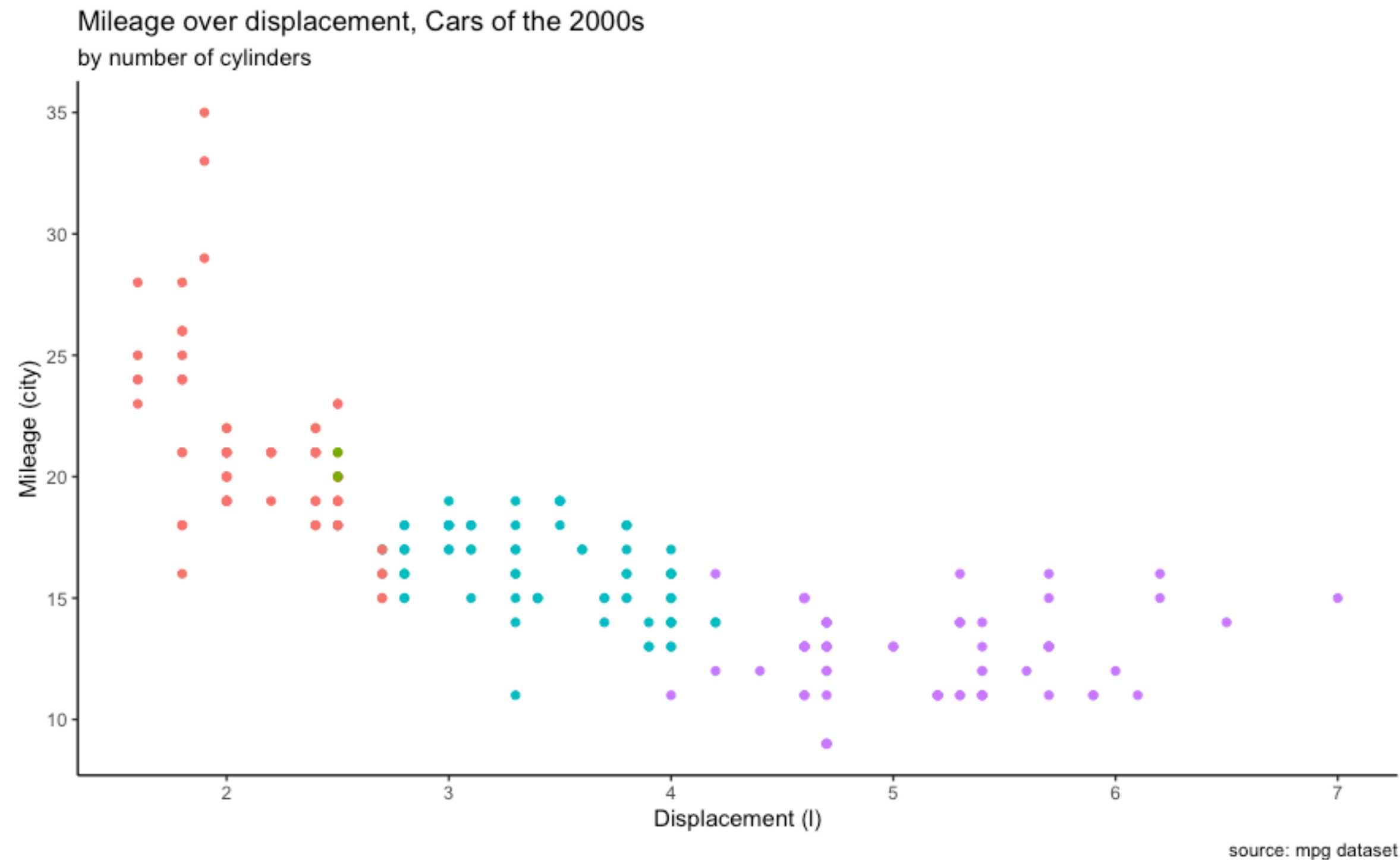
50 ways to show your data

- lollipop charts [ggalt]
alternative to charts with lots of nearly-similar-size bars,
on screen: reduce moiré, on paper: reduce waste of ink



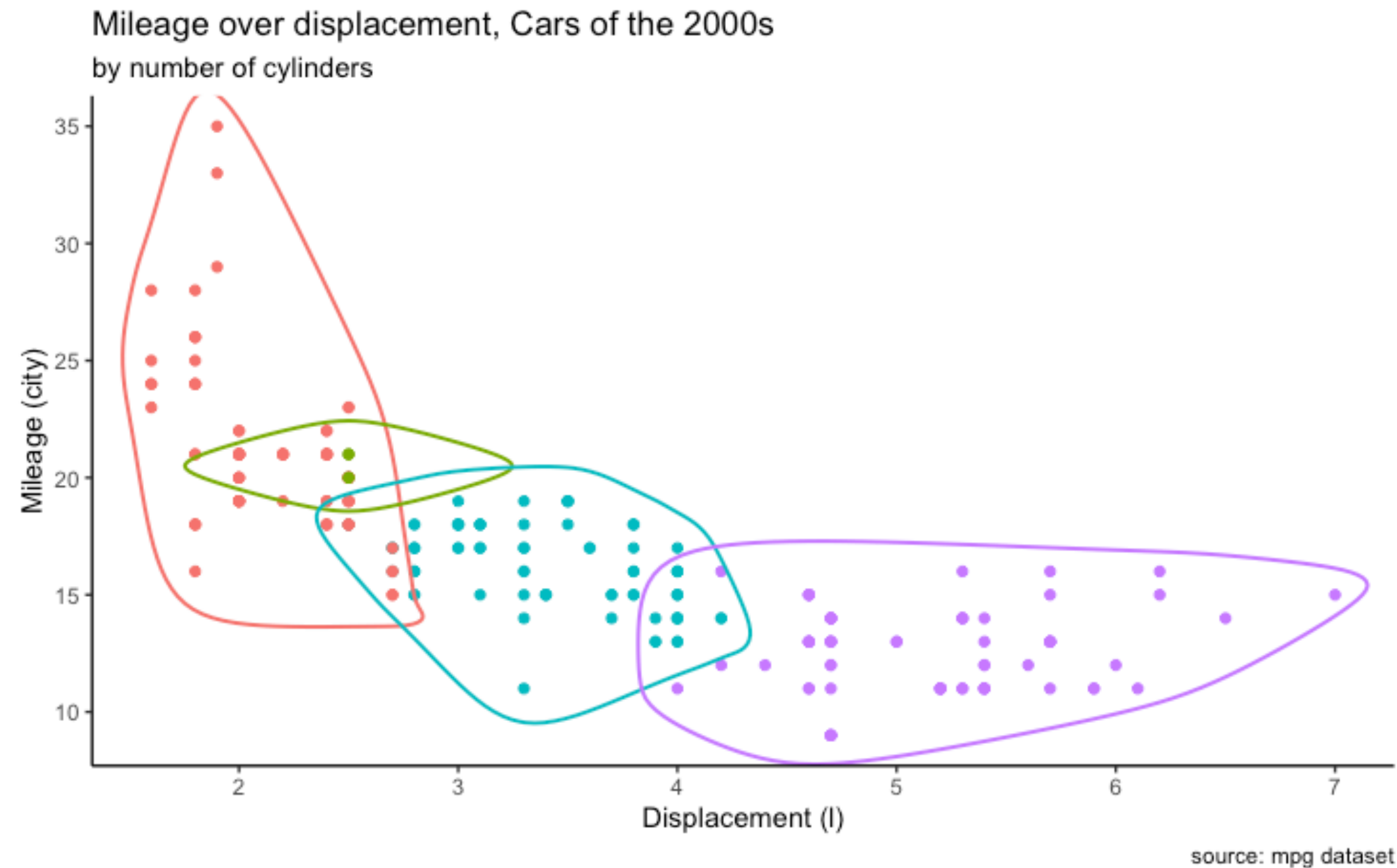
50 ways to show your data

- encircling areas [ggalt]
draw attention to certain regions or groups of points,
visualize clustering



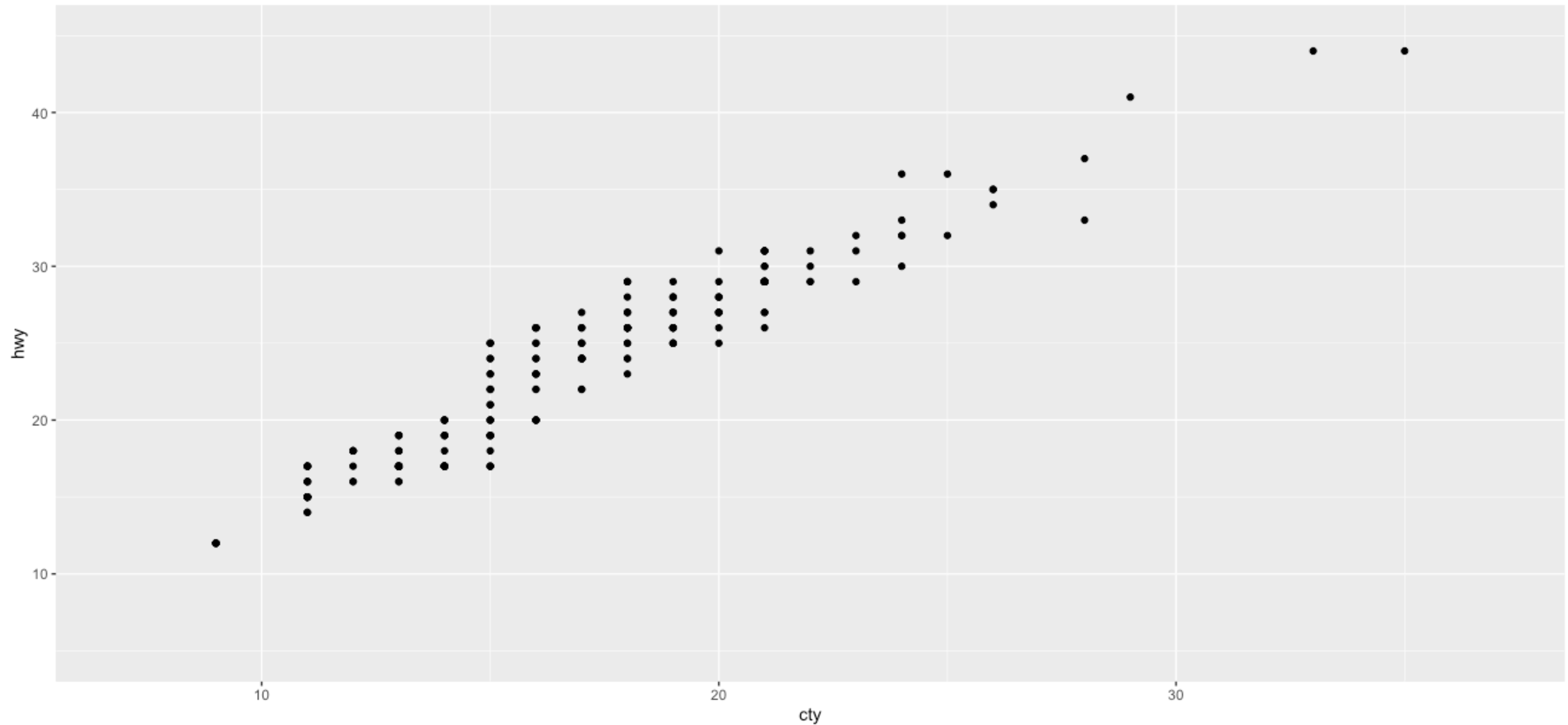
50 ways to show your data

- encircling areas [ggalt]
draw attention to certain regions or groups of points,
visualize clustering



50 ways to show your data

Scatterplot with 234 data points
mpg: city vs highway mileage

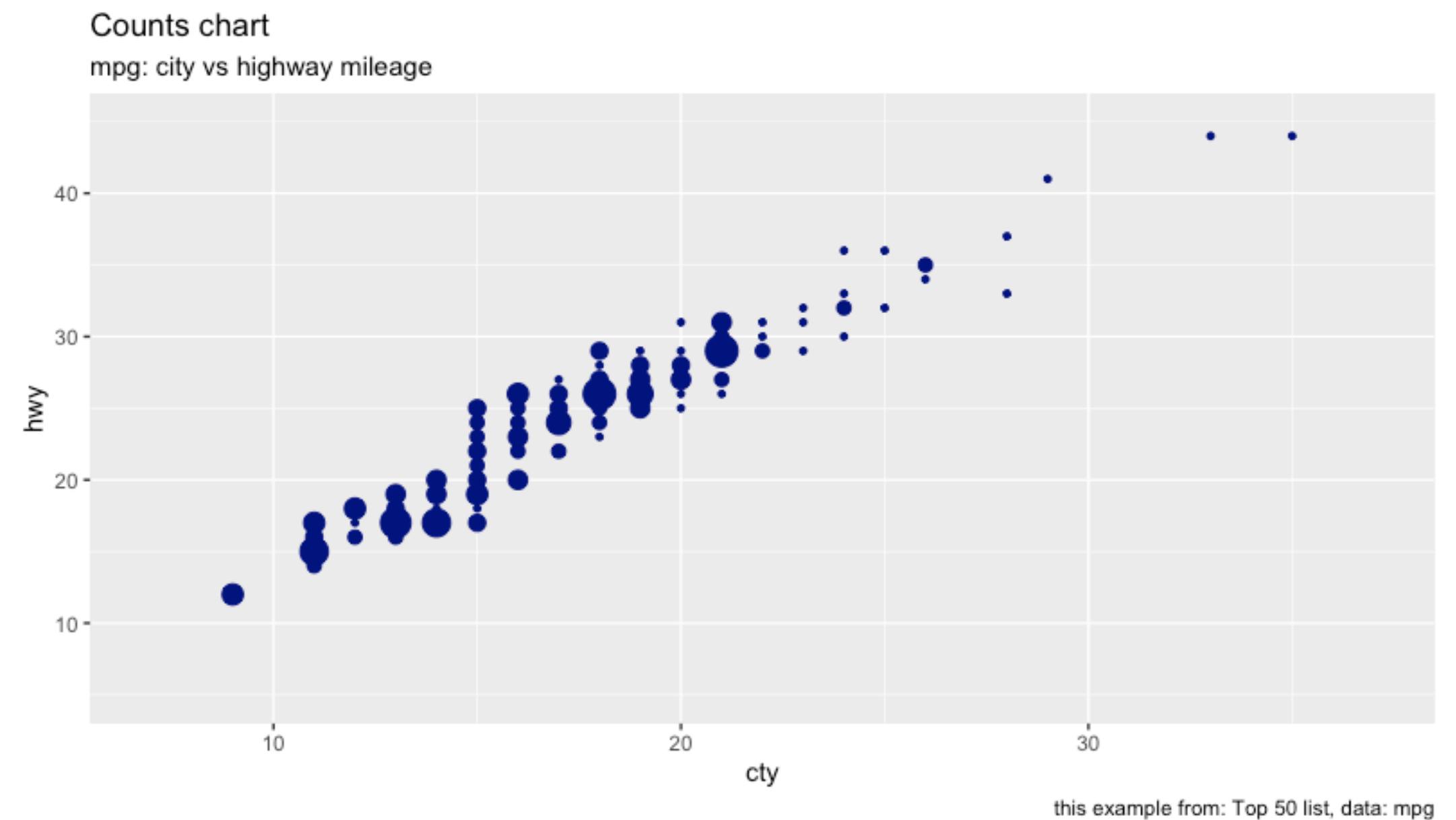
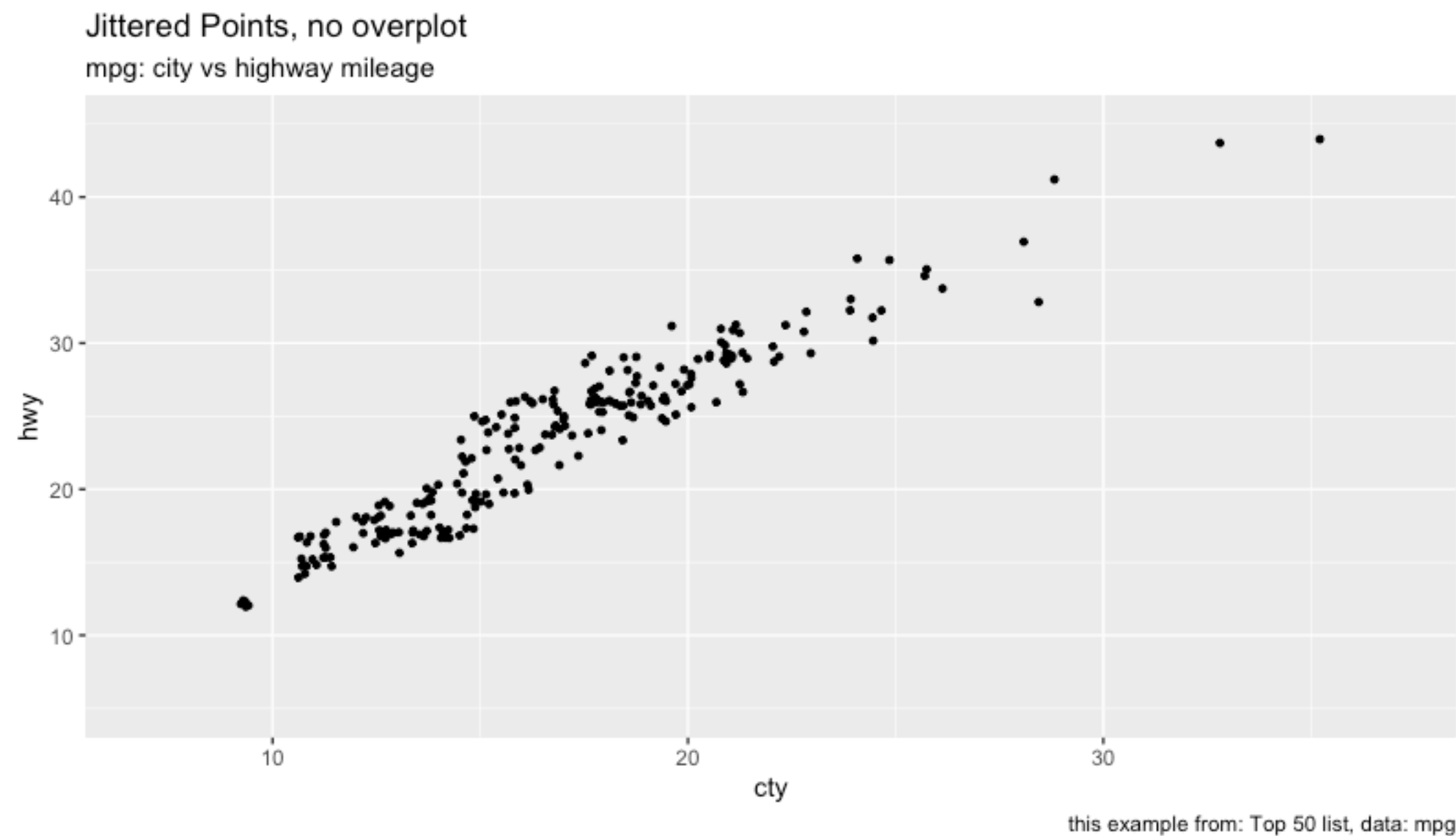


this example from: Top 50 list, data: mpg



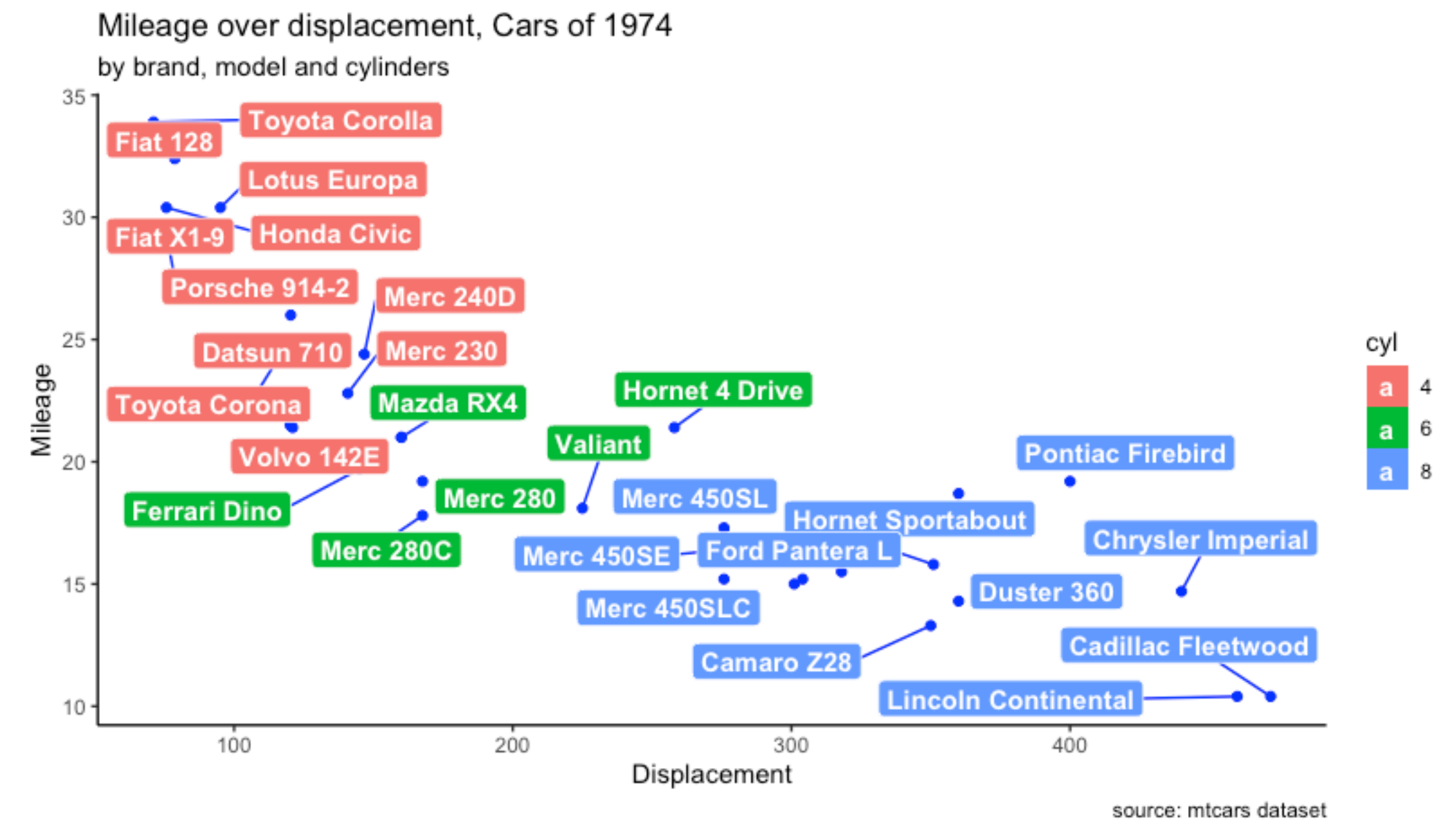
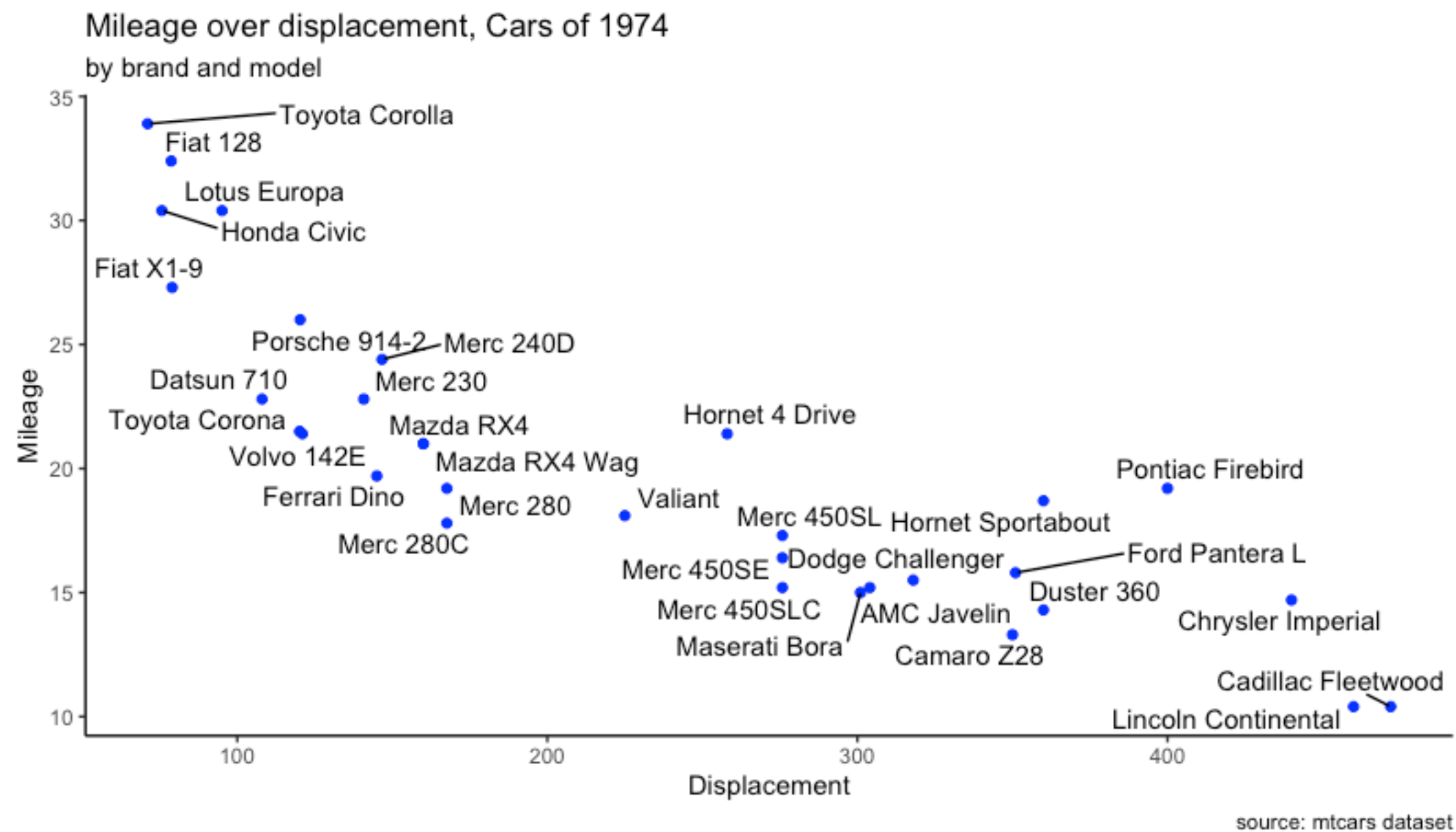
50 ways to show your data

- jitter plots / counts charts
show information otherwise hidden by
overlapping points



50 ways to show your data

- avoiding overlap of text labels [ggrepel]
avoid overlap of text labels



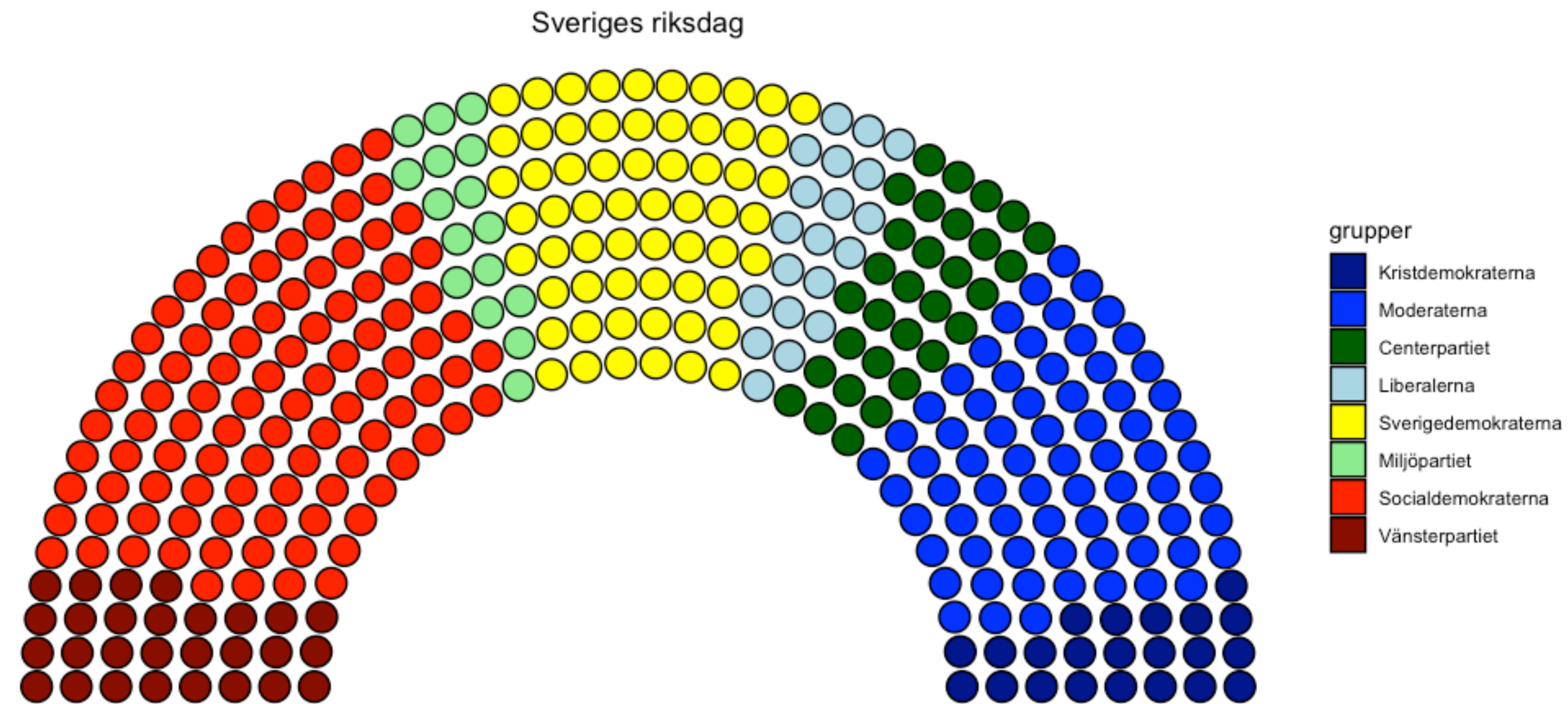
50 ways to show your data

- parliament diagrams [ggpol]
by seat number or vote distribution



50 ways to show your data

- parliament diagrams [ggpol]
by seat number or vote distribution

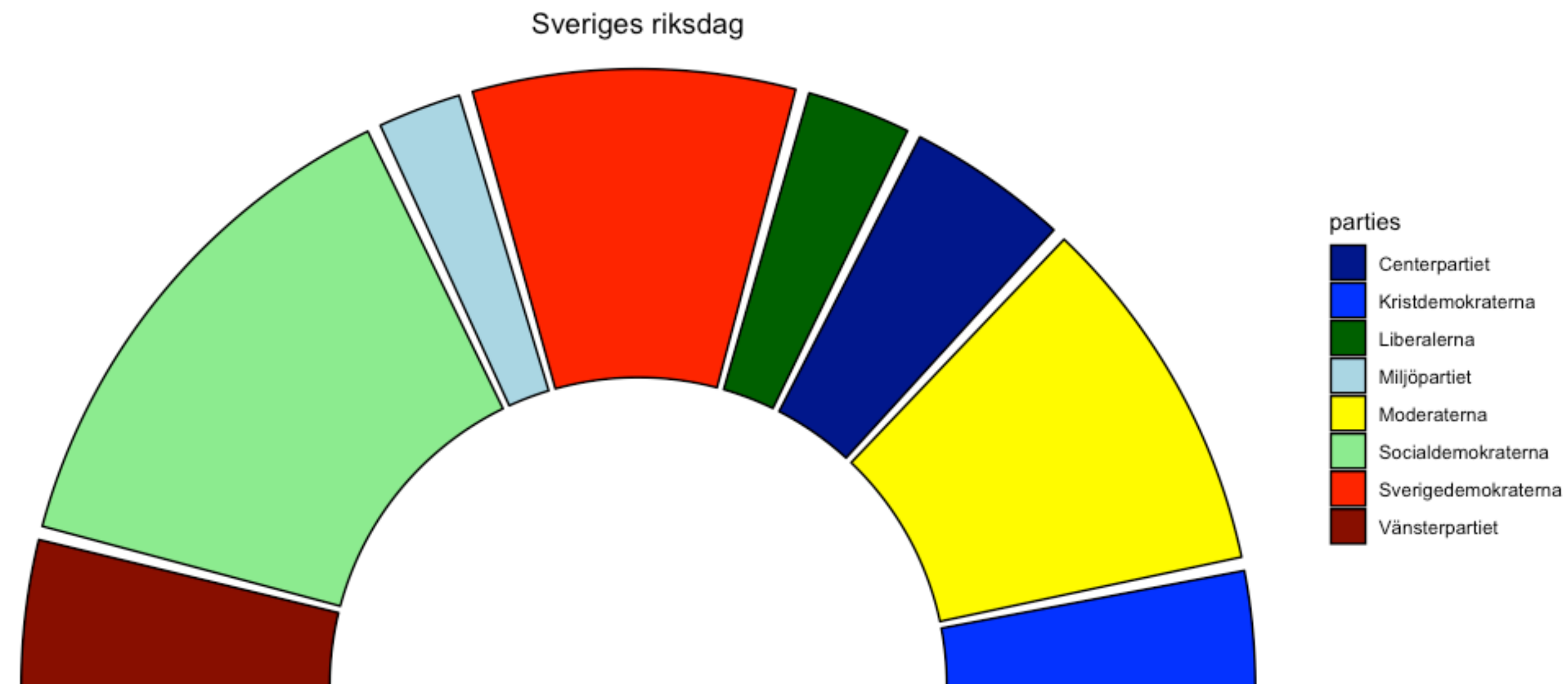


Source: Riksdagen 09/21

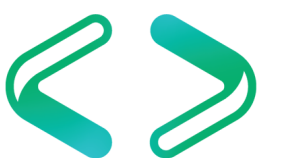


50 ways to show your data

- parliament diagrams [ggpol]
by seat number or vote distribution

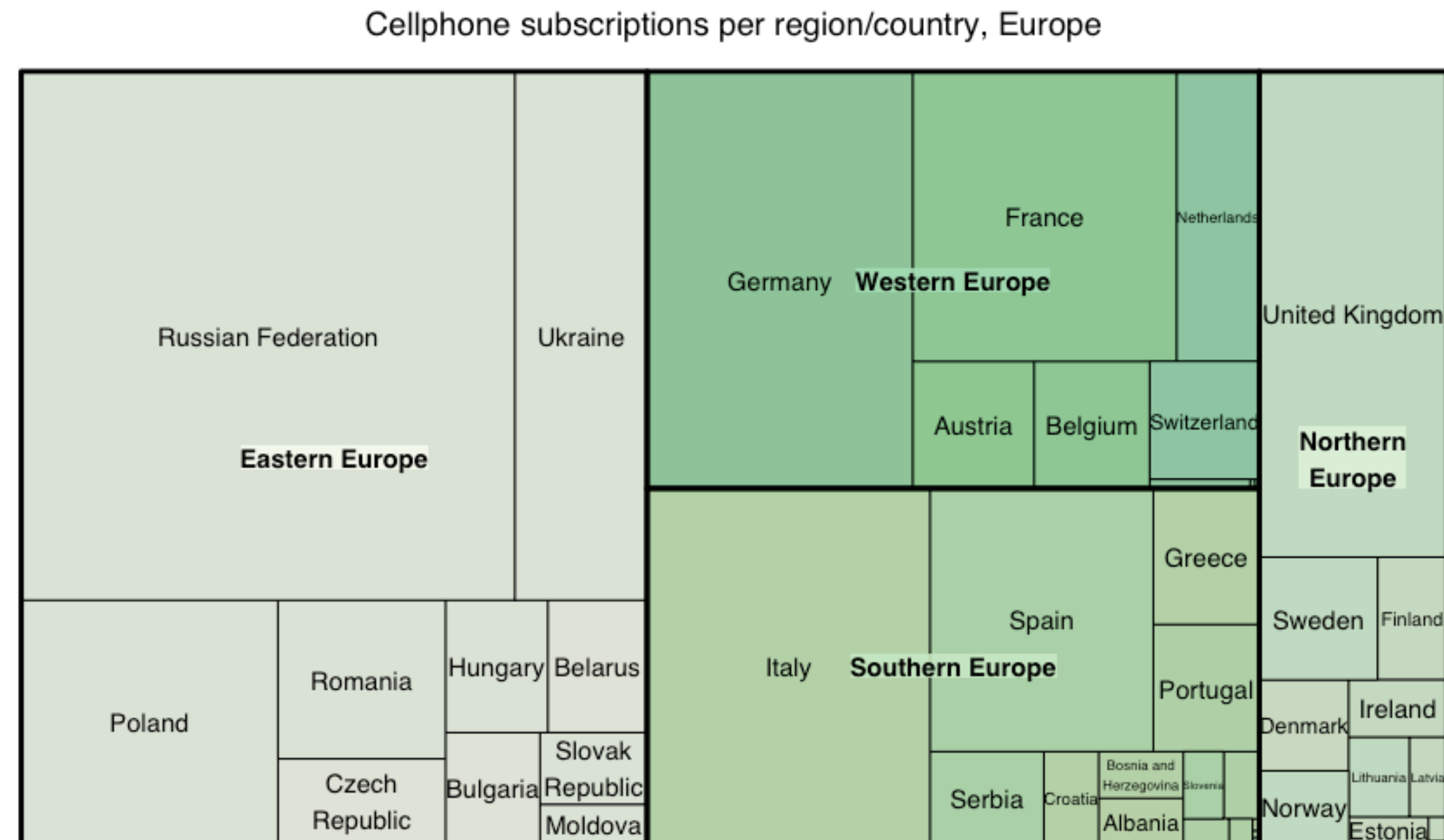


Source: Riksdagen 09/21



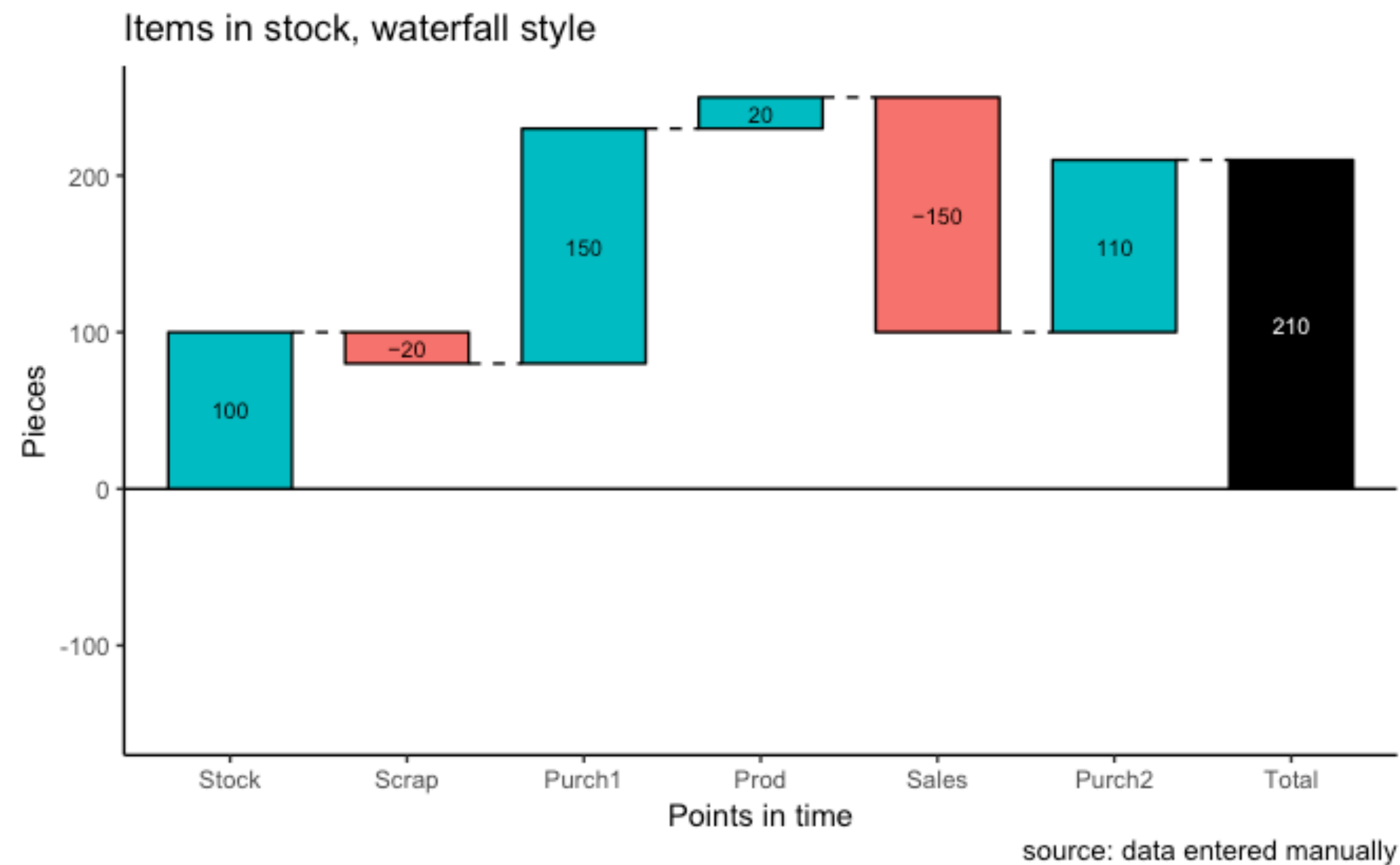
50 ways to show your data

- tree maps [treemap]
display hierarchical data as nested rectangles



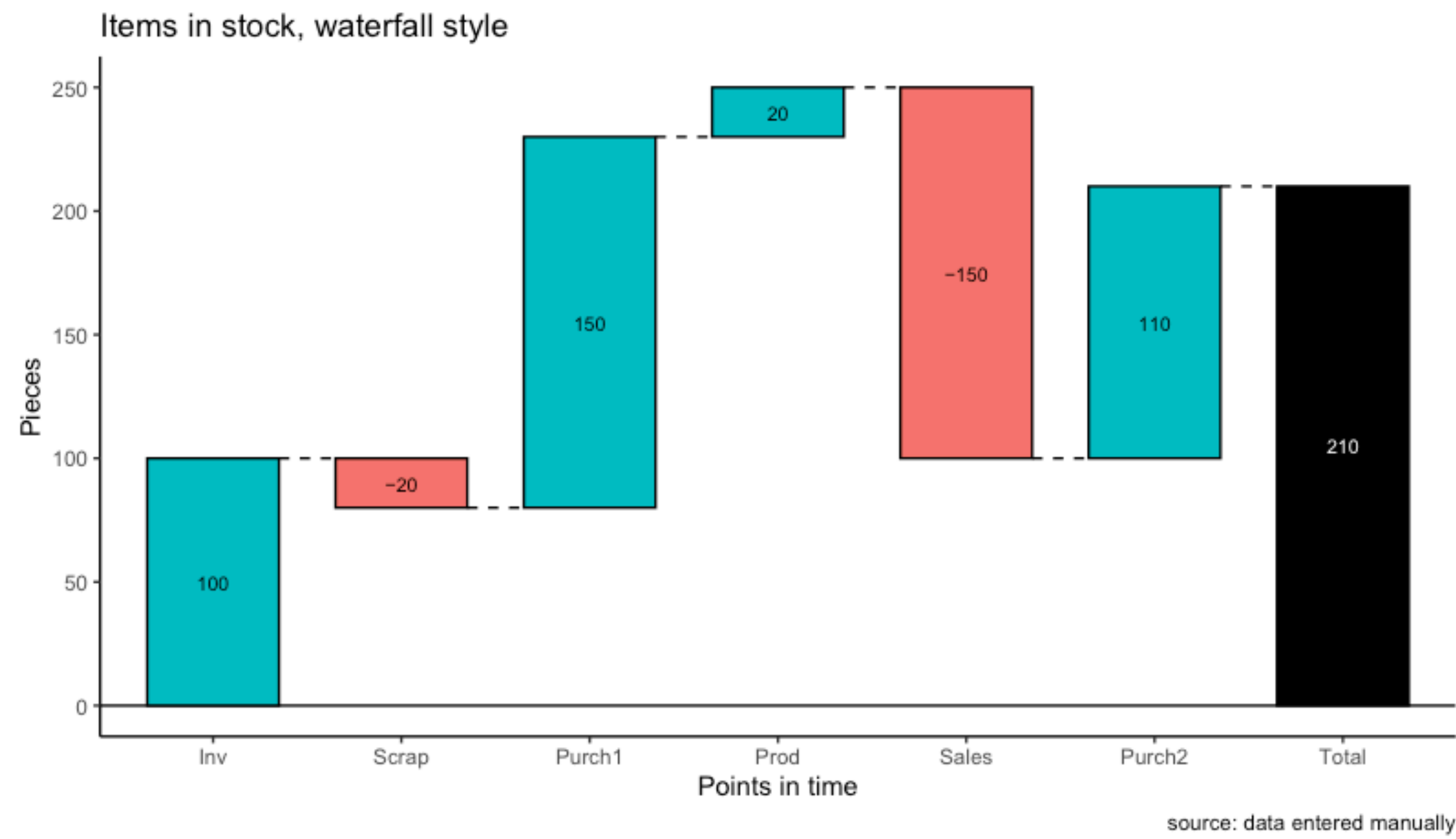
50 ways to show your data

- waterfall diagrams [waterfall(s)]
show up- and down development of a single measure



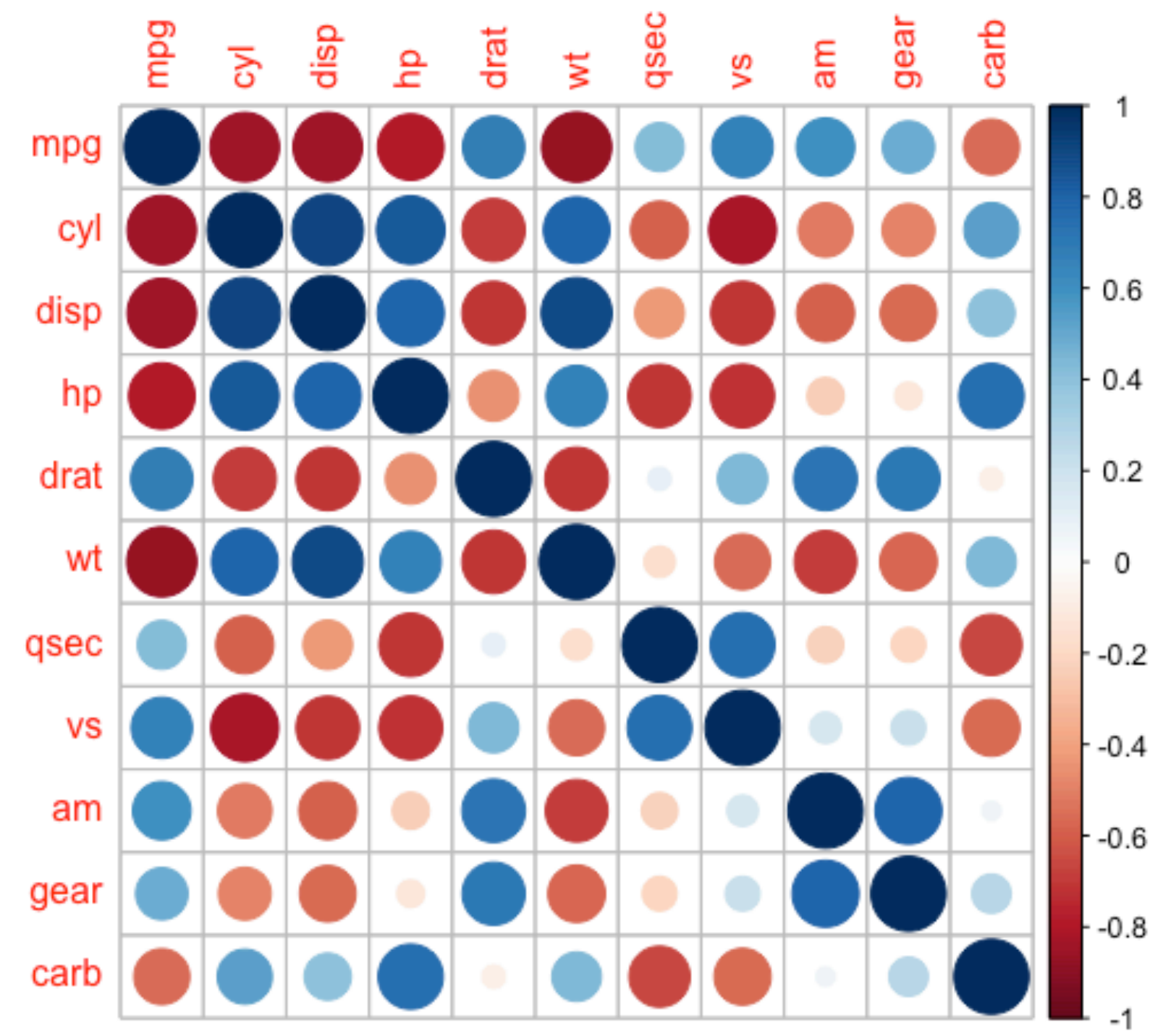
50 ways to show your data

- waterfall diagrams [waterfall(s)]
show up- and down development of a single measure



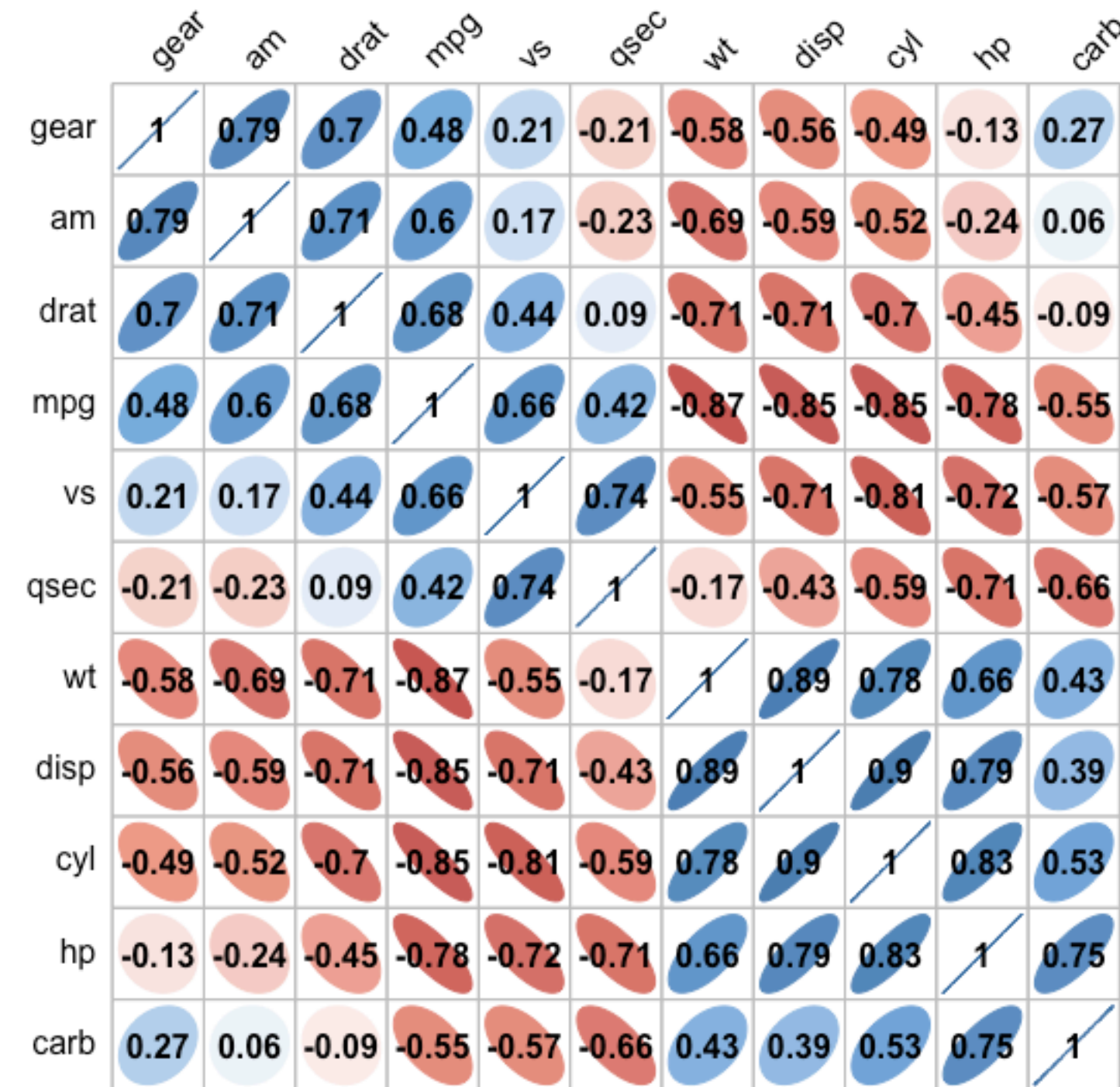
50 ways to show your data

- correlograms [(gg)corrplot]
show correlation between several variables at once



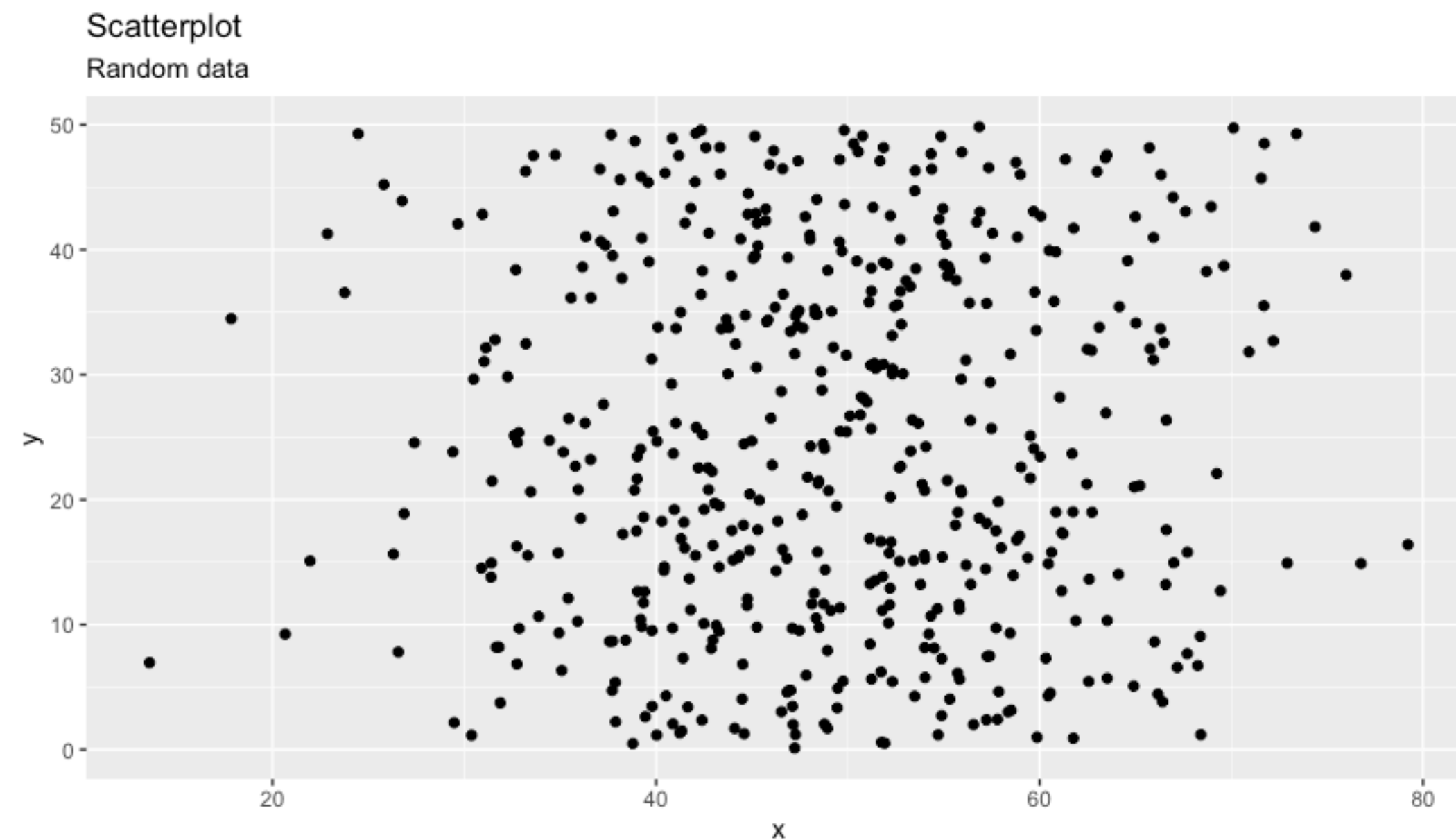
50 ways to show your data

- correlograms [(gg)corrplot]
show correlation between several variables at once



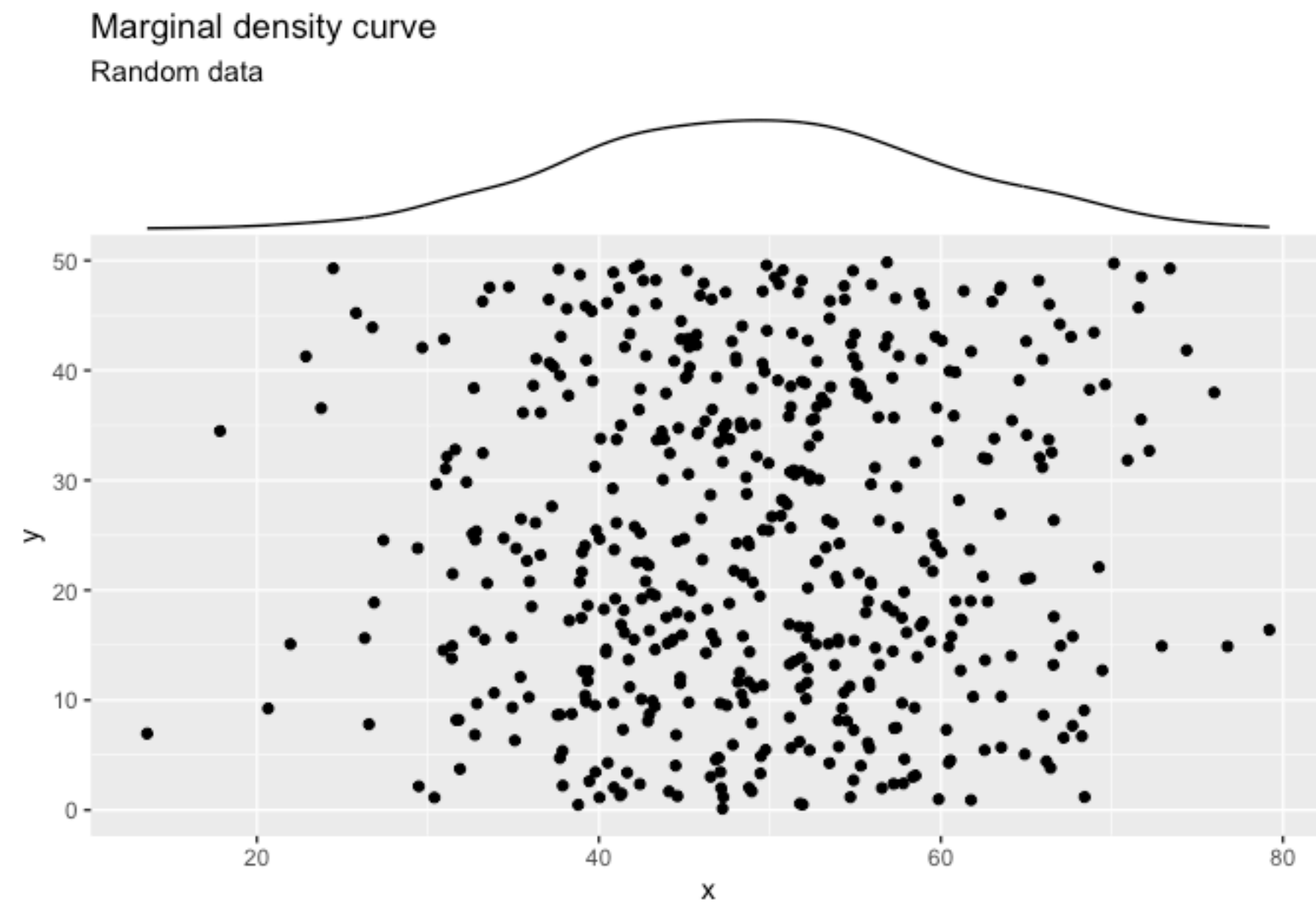
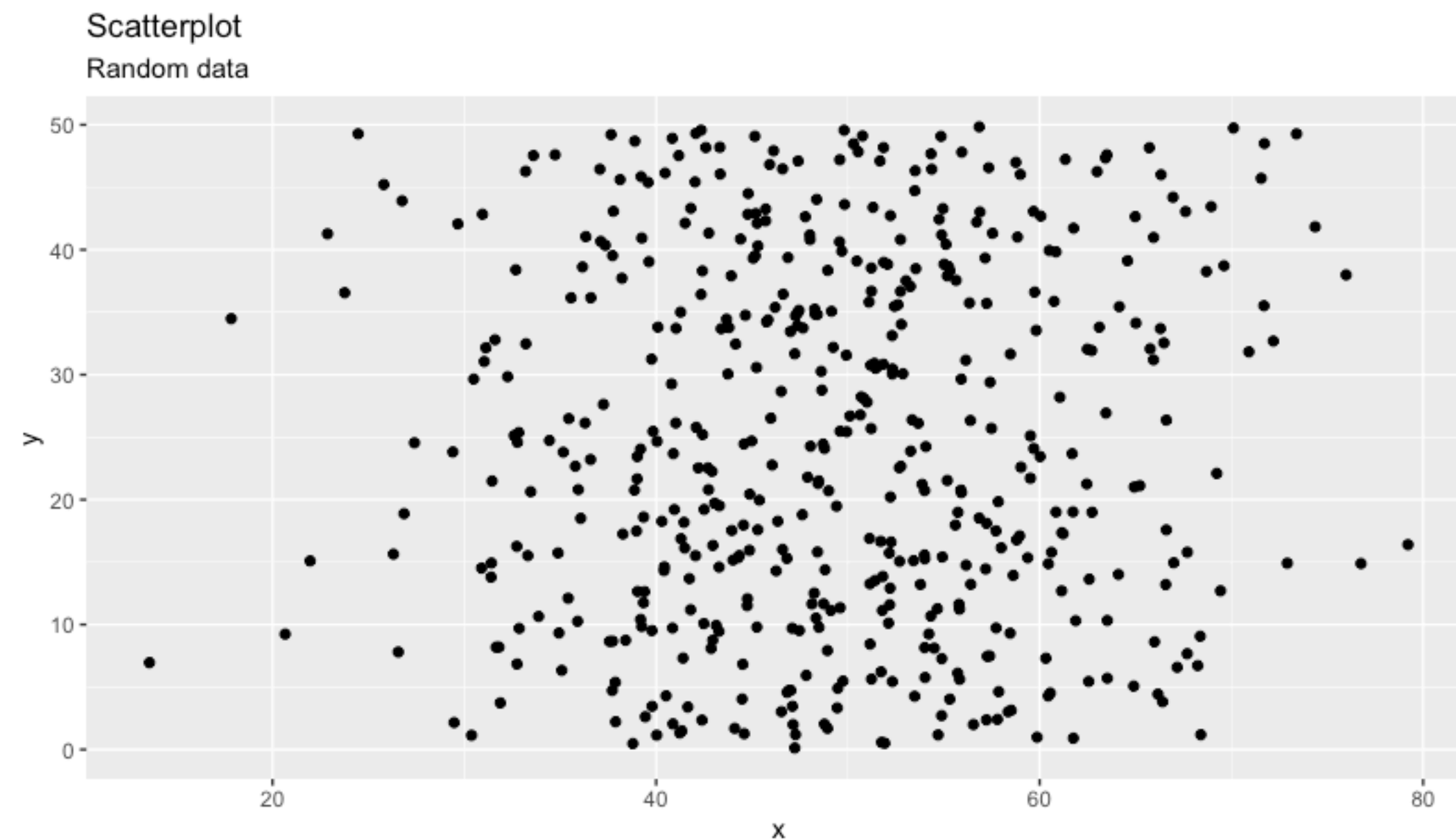
50 ways to show your data

- marginal histograms and boxplots [ggExtra]
show distribution details for the dimensions of
a scatter plot



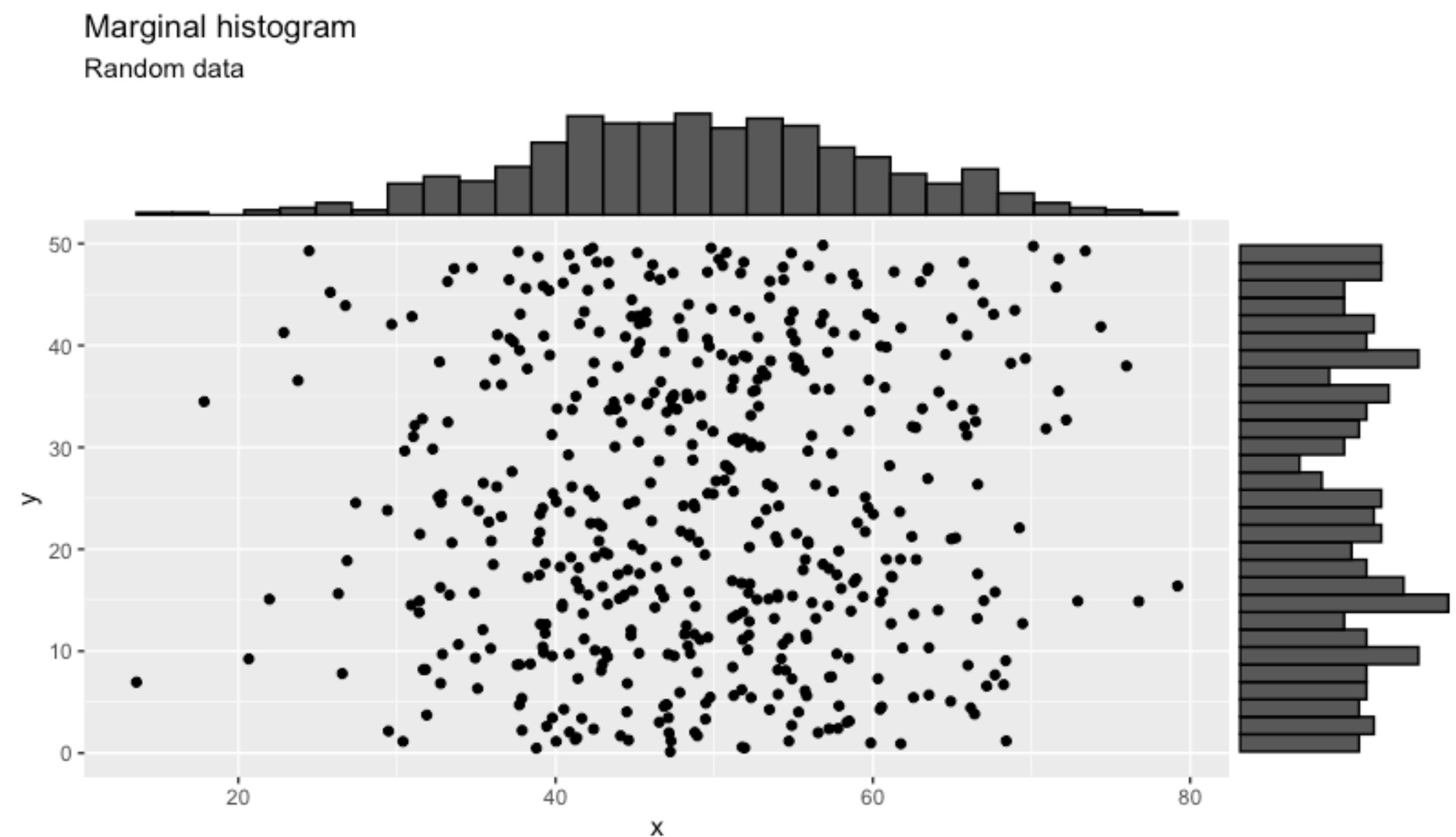
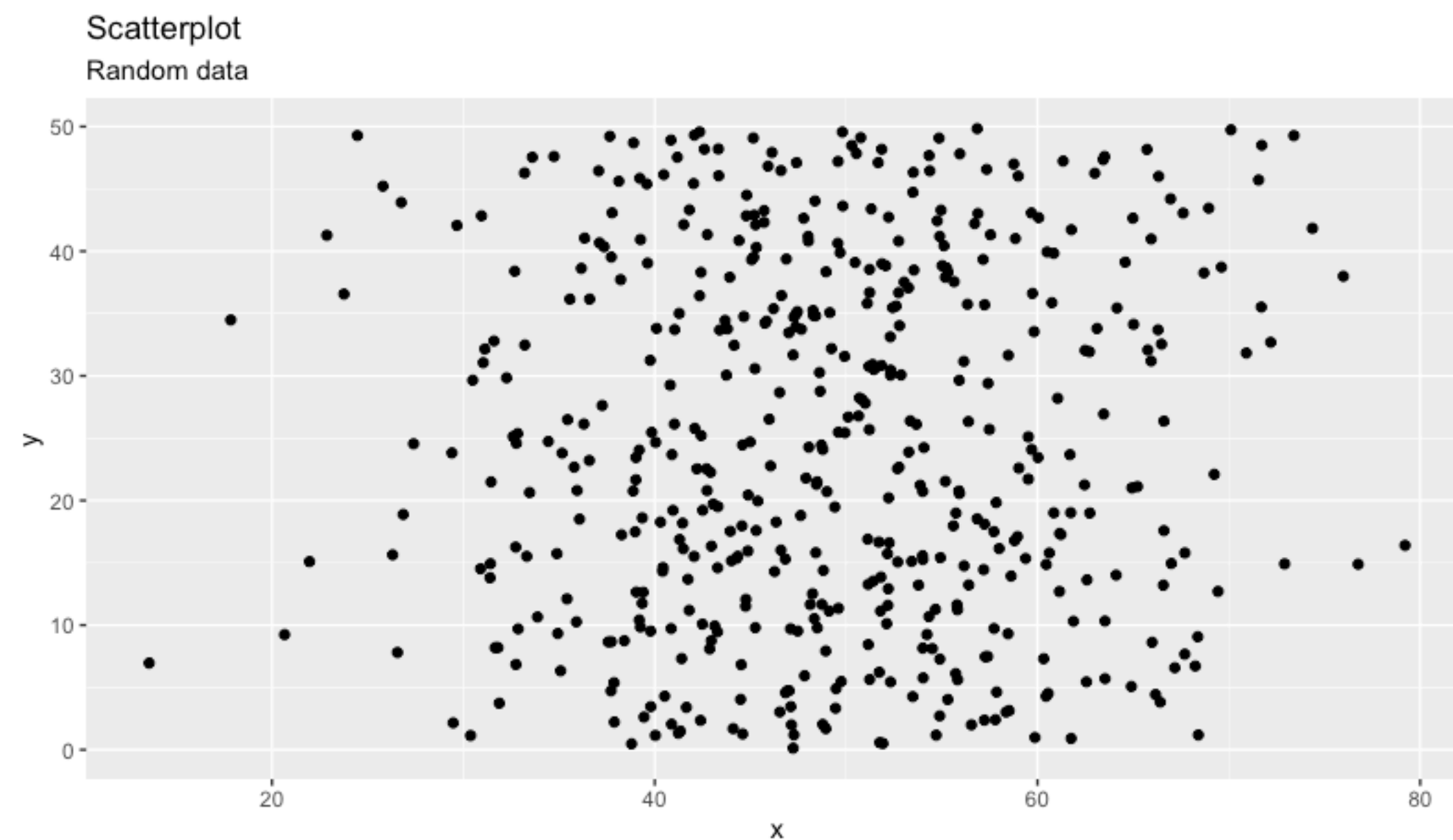
50 ways to show your data

- marginal histograms and boxplots [ggExtra]
show distribution details for the dimensions of
a scatter plot



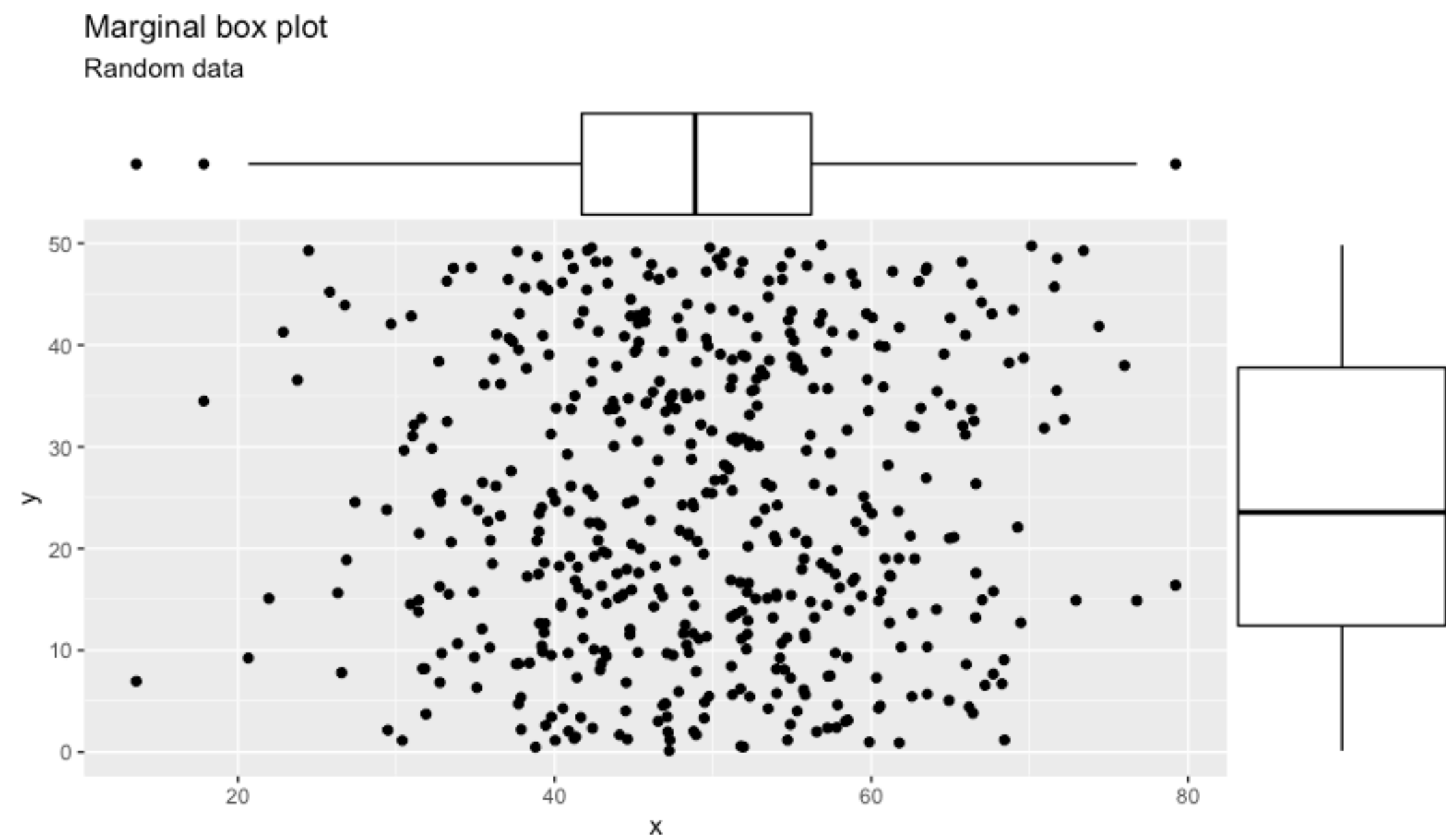
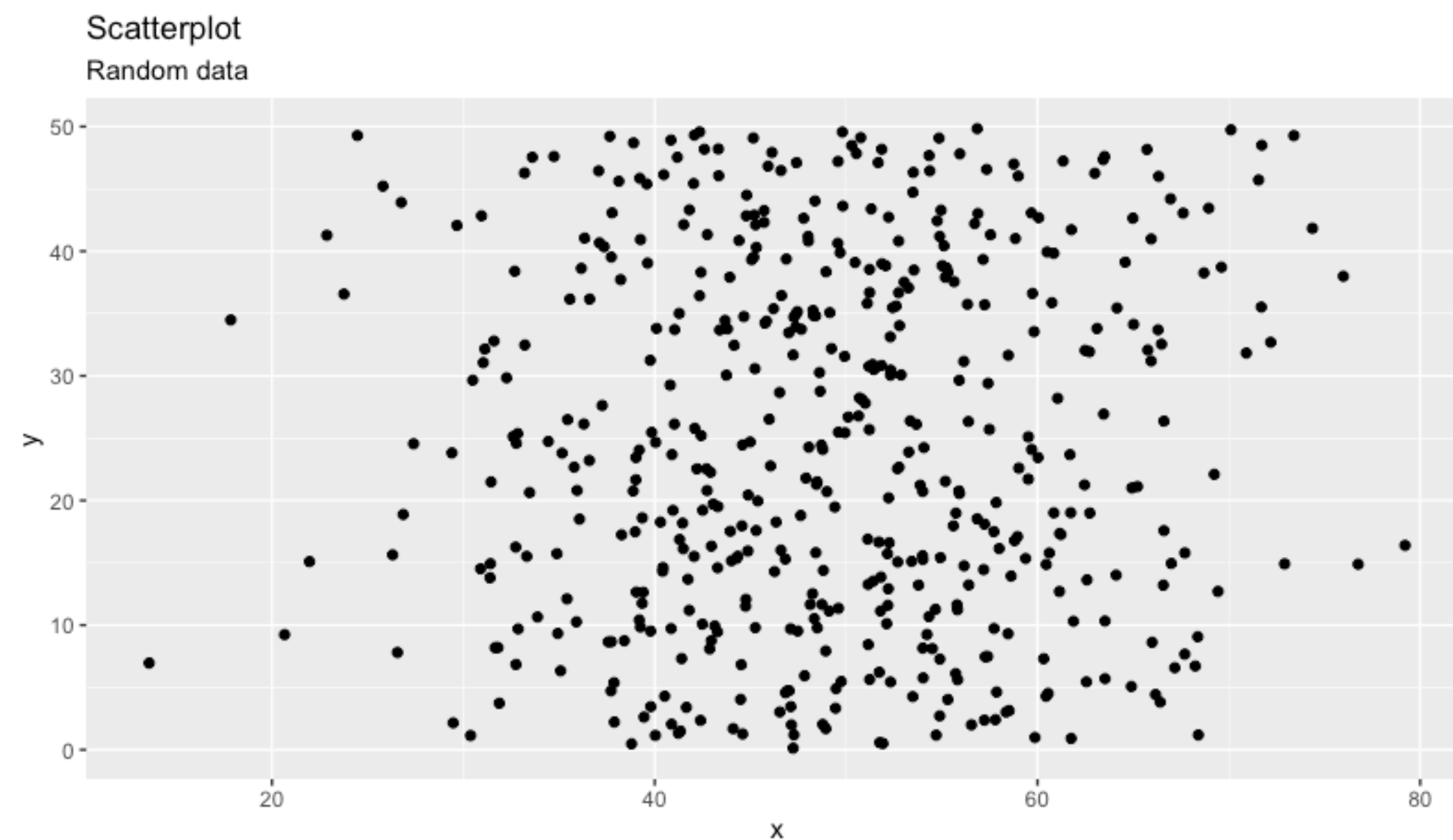
50 ways to show your data

- marginal histograms and boxplots [ggExtra]
show distribution details for the dimensions of
a scatter plot



50 ways to show your data

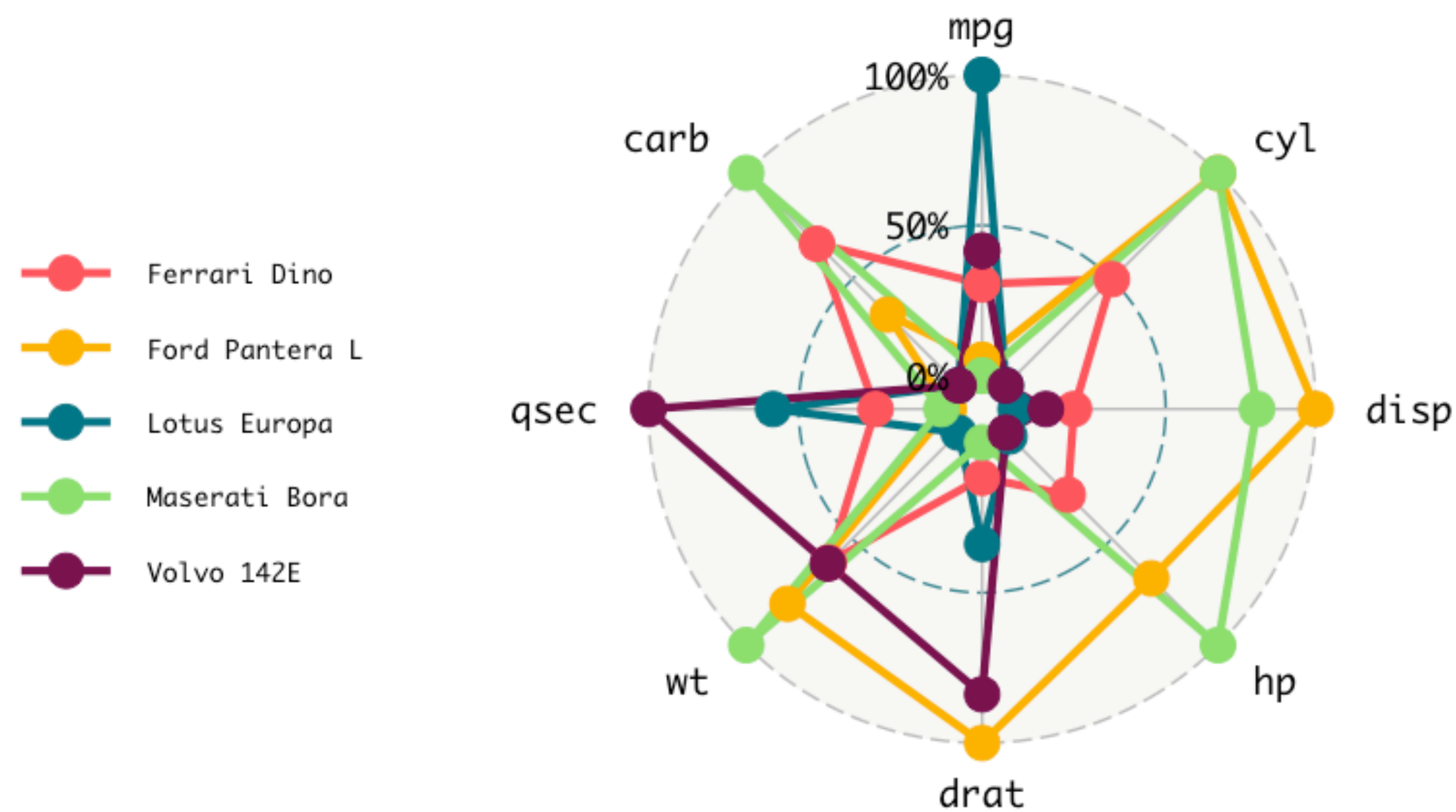
- marginal histograms and boxplots [ggExtra]
show distribution details for the dimensions of
a scatter plot



50 ways to show your data

- radar charts [ggradar]
compare multiple measures for few items

Compare multiple properties



50 ways to show your data

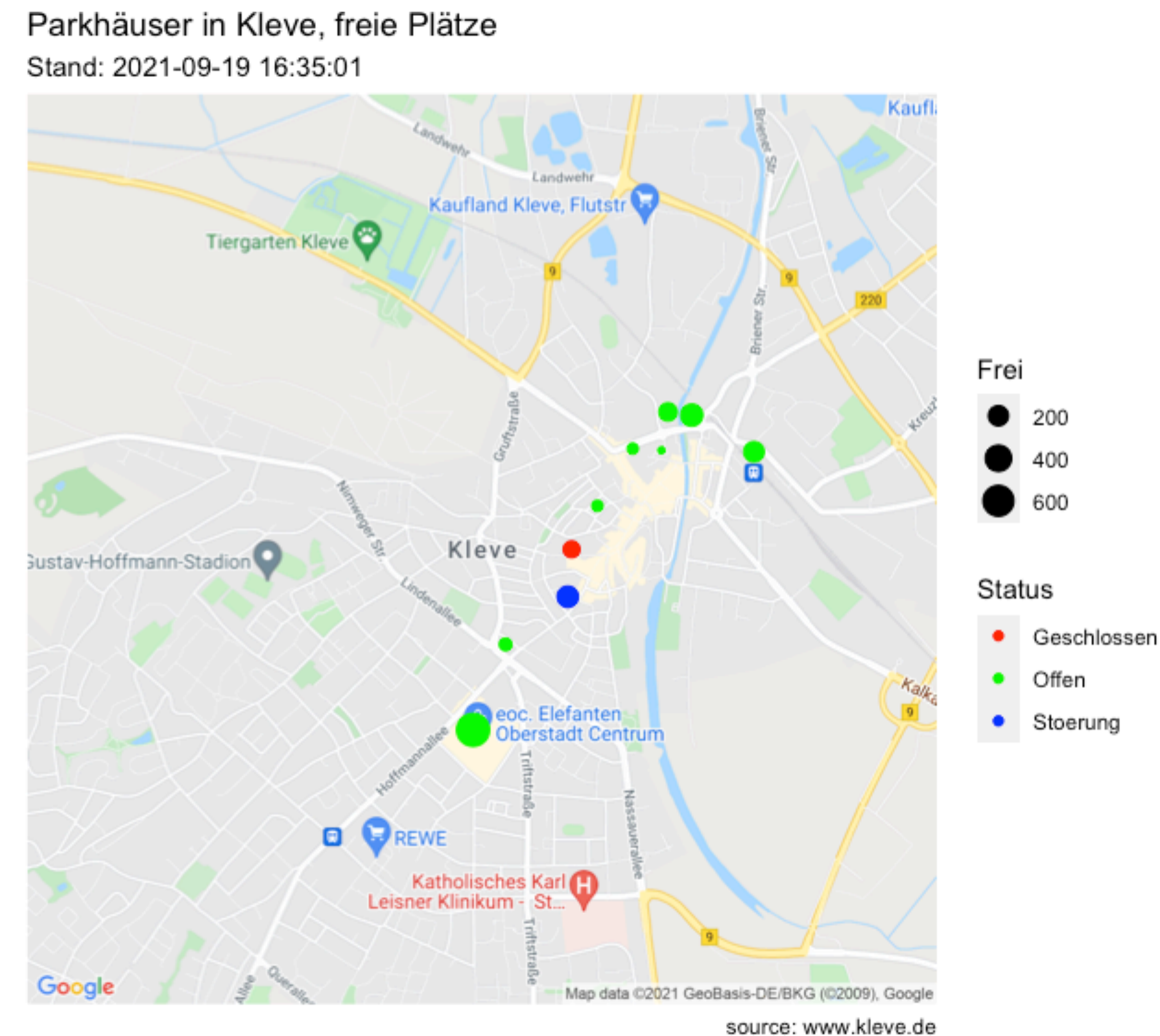
- radar charts [ggradar]
compare multiple measures for few items

```
mtcars %>%  
  rownames_to_column(var = "group") %>%  
  tail(5) %>% mutate_at(vars(-group), list(rescale)) %>%  
  select(1:8,12) -> mtcars_radar  
  
ggradar(mtcars_radar, font.radar = 'Ayuthaya',  
        plot.title = 'Compare multiple properties',  
        legend.text.size = 10)
```



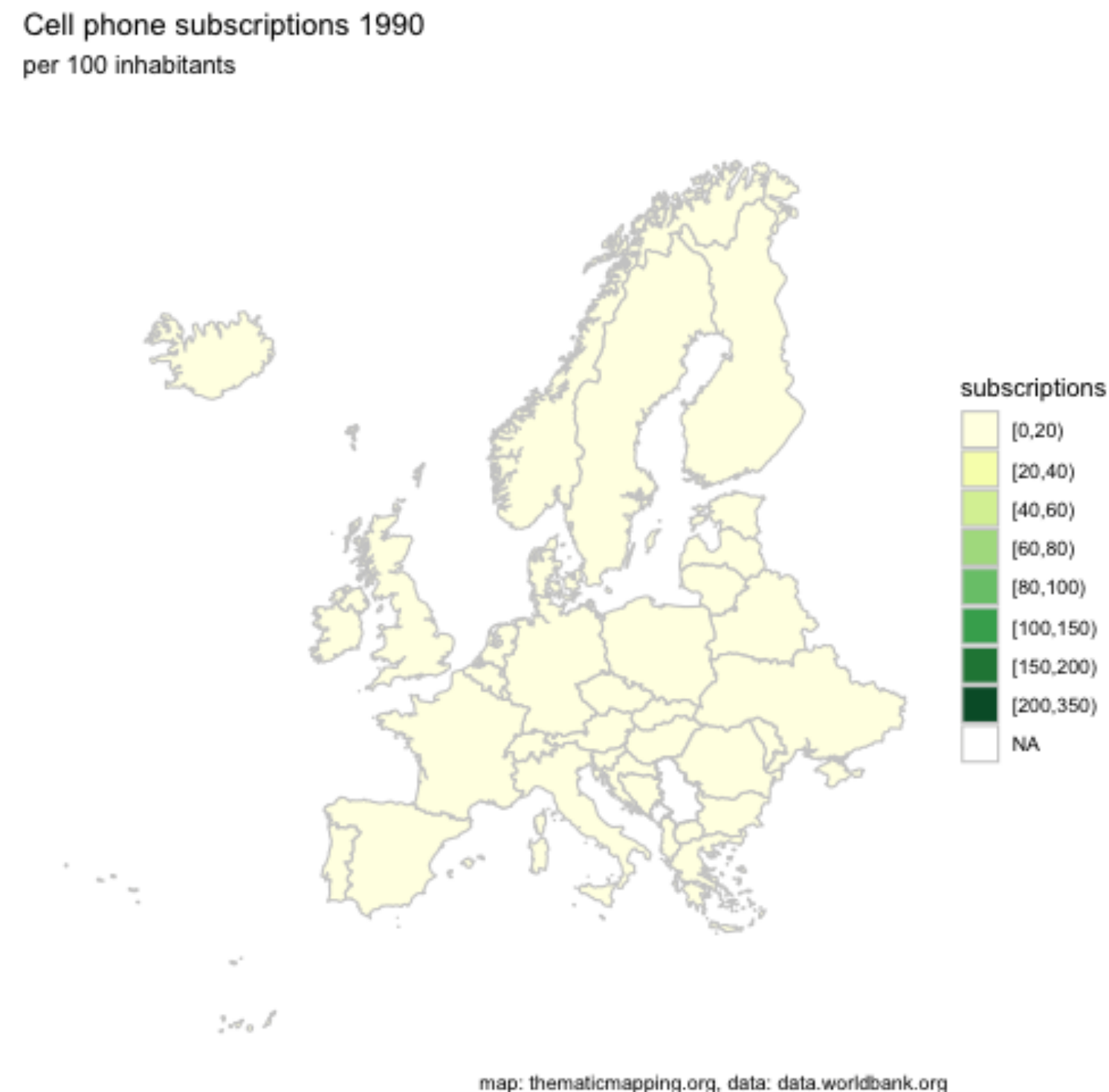
50 ways to show your data

- maps [ggmap]
include geospatial information



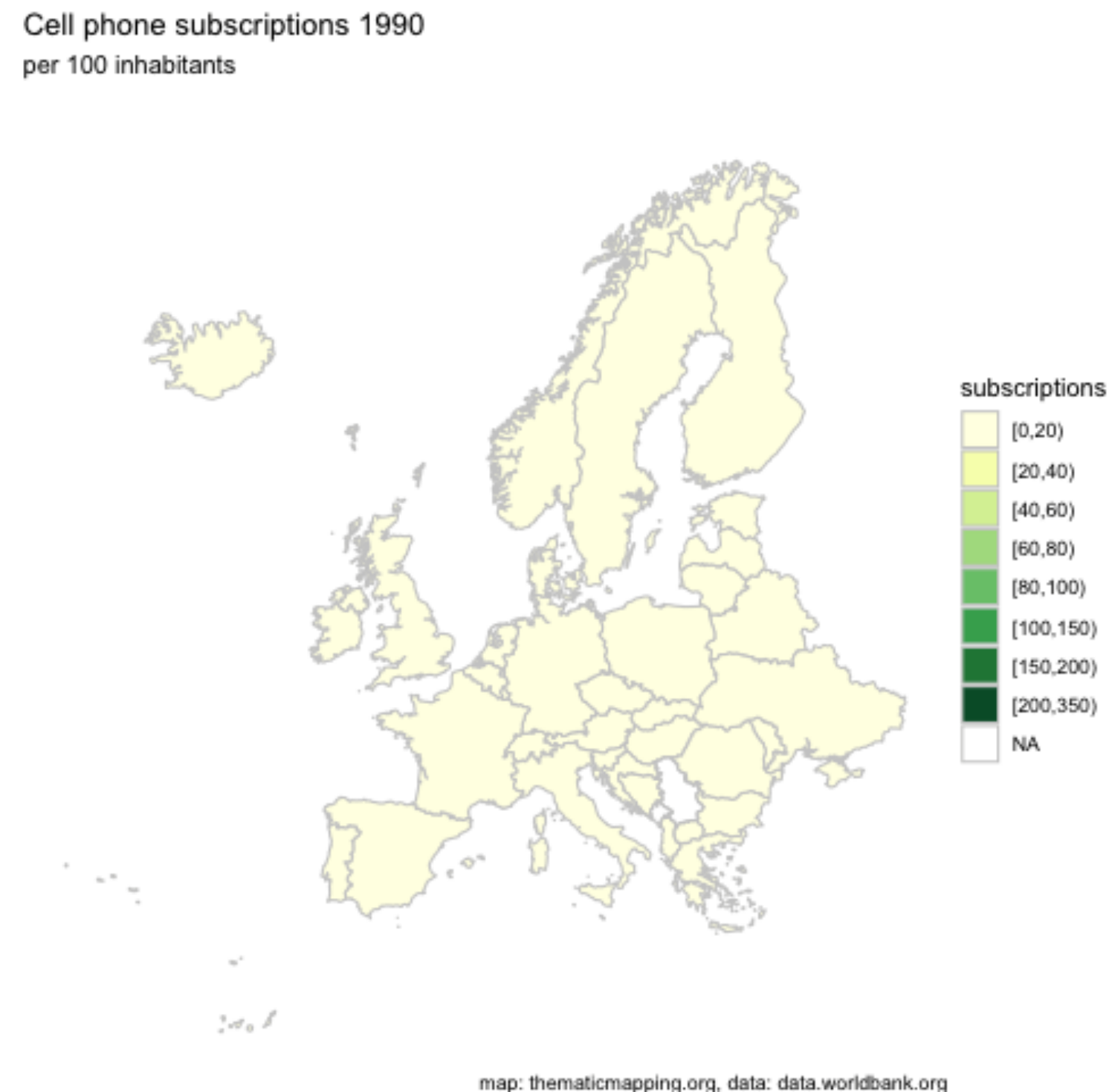
50 ways to show your data

- animated plots [gganimate]
automatically show development over time



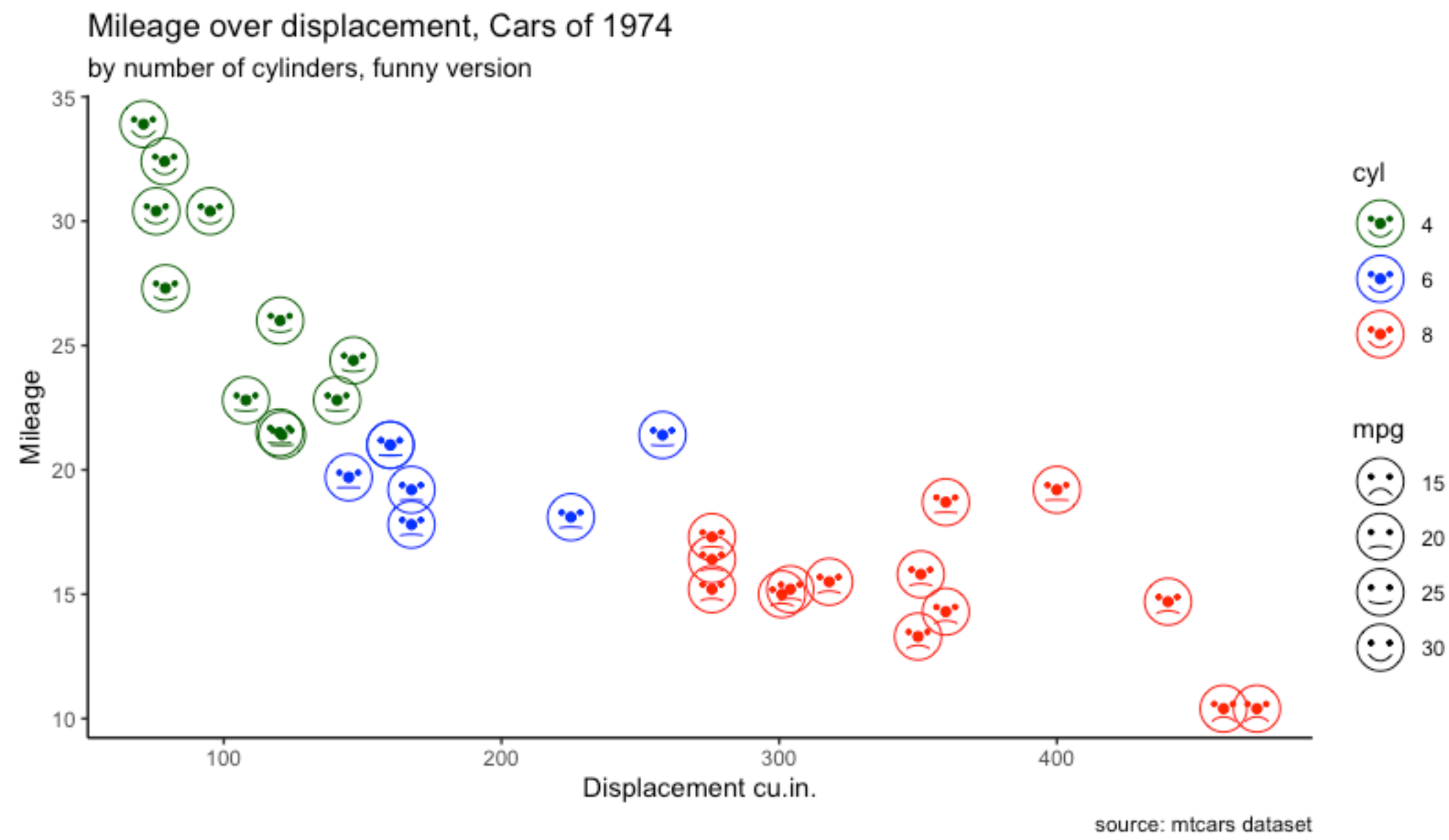
50 ways to show your data

- animated plots [gganimate]
automatically show development over time



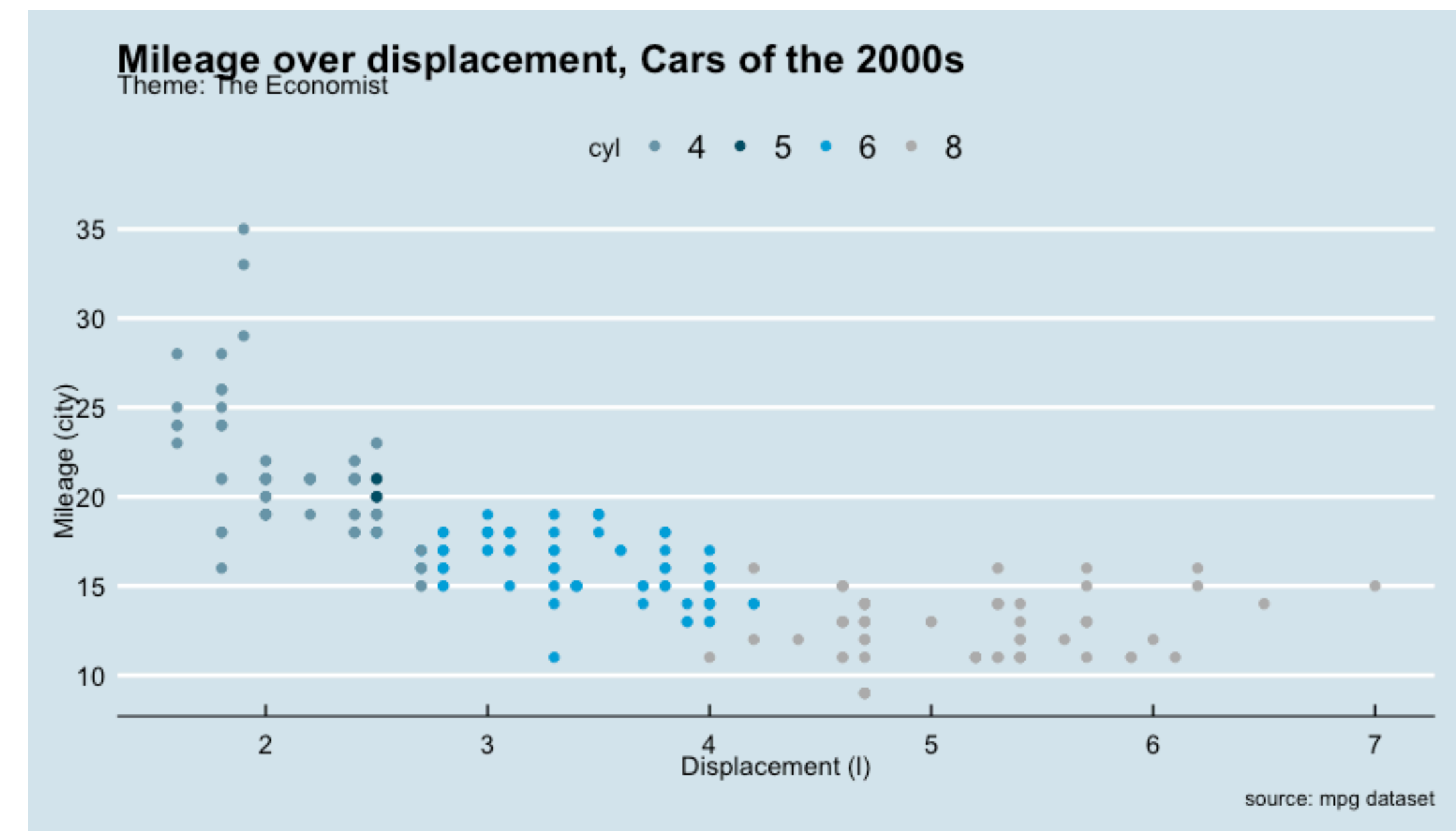
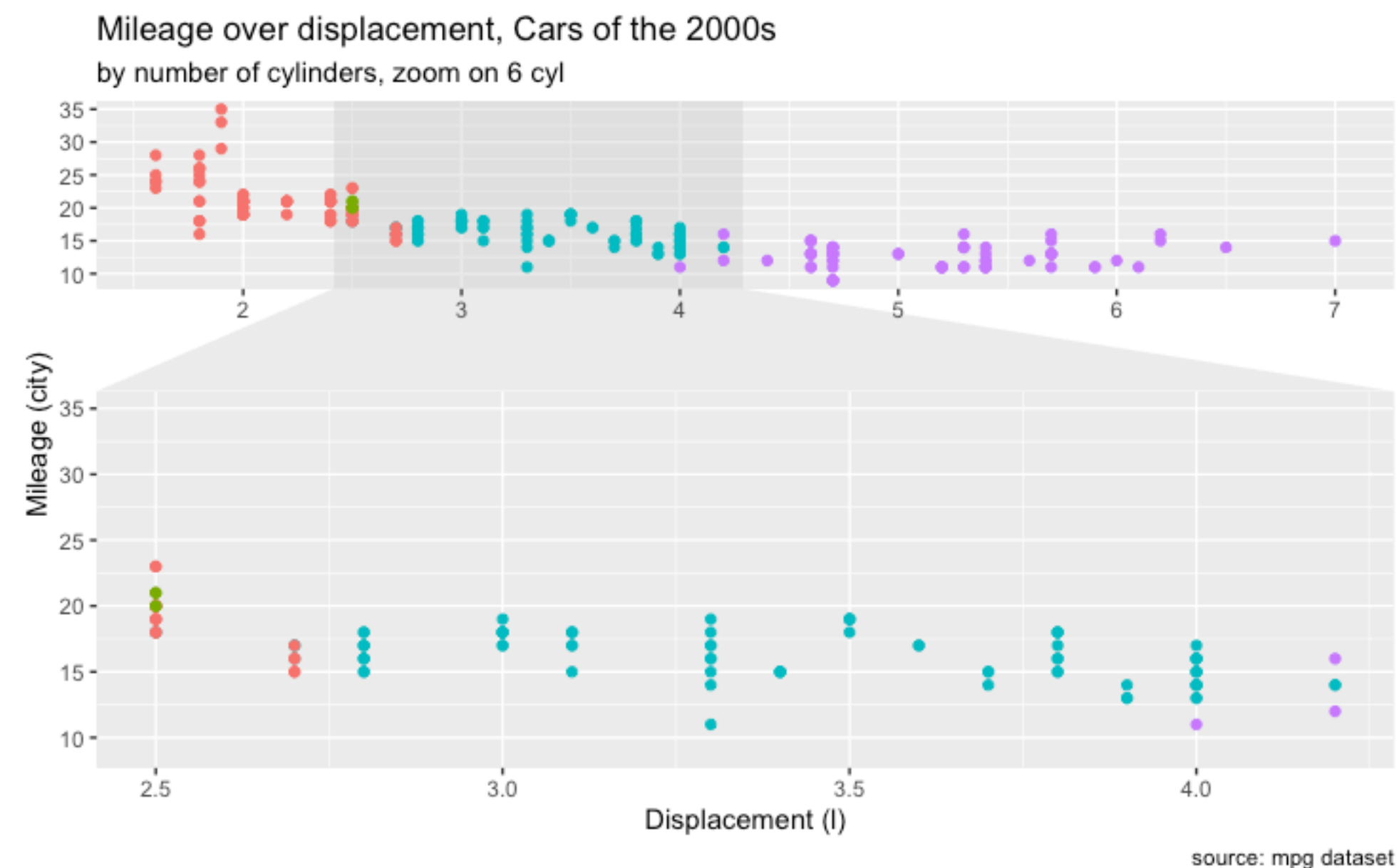
50 ways to show your data

- Chernoff faces [ggChernoff]
everyone likes smilies, don't they?



50 ways to show your data

- More extras:
facet zooming [ggforce], diverse themes [ggthemes, ggTech], interactive: tooltips [ggiraph]



Round-up / conclusions



Round-up / conclusions

- Even with open source products like R, a lot is possible in terms of analysis and compelling visualizations



Round-up / conclusions

- Even with open source products like R, a lot is possible in terms of analysis and compelling visualizations
- Applying the Tidyverse philosophy and tools, ggplot2 and its extensions, graphs can be constructed in a very concise manner, with very small effort of coding



Round-up / conclusions

- Even with open source products like R, a lot is possible in terms of analysis and compelling visualizations
- Applying the Tidyverse philosophy and tools, ggplot2 and its extensions, graphs can be constructed in a very concise manner, with very small effort of coding
- But - even in this context - sometimes less is more: Don't overdo/overload, carefully choose colours/shapes, consider the occasion of your presentation (Chernoff faces are not suitable for every audience)



Credits & resources / inspired by:

- Get base R and packages: <https://cran.r-project.org/index.html>
- Get RStudio: <https://www.rstudio.com/products/rstudio/download/>
- Selva Prabhakaran's [Top 50 ggplot2 Visualizations Master List](#)
(use under [Creative commons license](#), referred to as „Top 50 list“)
- A list of ggplot2 extensions <https://exts.ggplot2.tidyverse.org/>
- The ggplot2 reference online <http://ggplot2.tidyverse.org/reference/>
- R Graphics Cookbook, by Winston Chang, O'Reilly, ISBN 978-1449316952
- R for Data Science, Hadley Wickham & Garrett Grolemund, O'Reilly, ISBN 978-1491910399, also online at <http://r4ds.had.co.nz>
- Data sources: www.formel1.de, data.worldbank.org, www.kleve.de



50 ways to show your data

Tack så mycket for your interest & keep in touch:

 @DerFredo <https://twitter.com/DerFredo>

 de.linkedin.com/in/derfredo

 www.xing.com/profile/Thomas_Huetter



This file and the demo script can be found at:
<https://j.mp/DerFredoDIS21>

