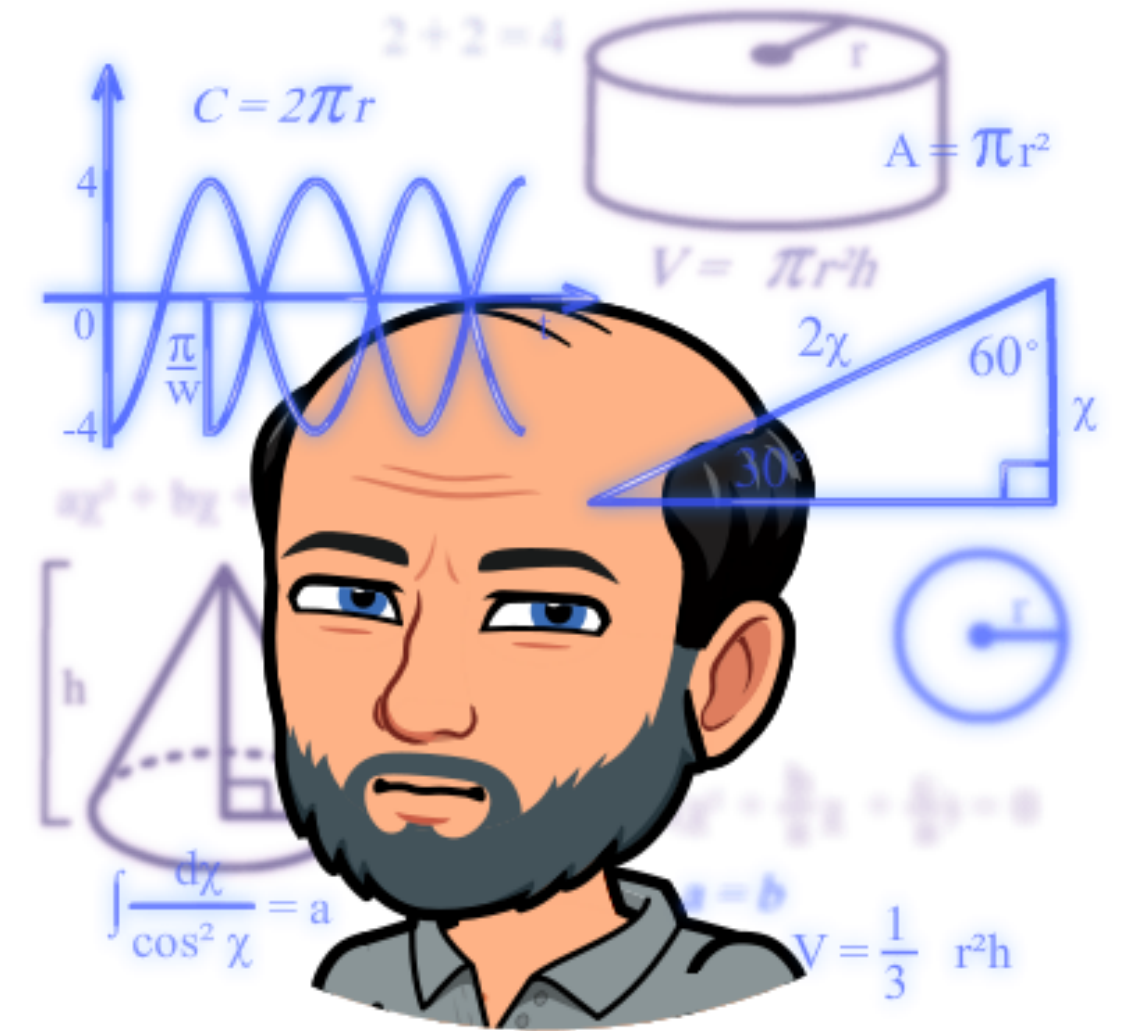


50 ways to show your data



Thomas Hütter

NDC London 2023, QEII Centre

50 ways to show your data

Thomas Hütter, Diplom-Betriebswirt

- Developer for ERP apps, SQL scripts and BI stuff
- Worked at consultancies, ISVs, end user companies
- 1995: SQL Server, 2001: Nav/BC, 2014: R, 2020: Power*
- Speaker at SQL / data / dev events around Europe

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

 de.linkedin.com/in/derfredo

 <https://techhub.social/@DerFredo>



Agenda

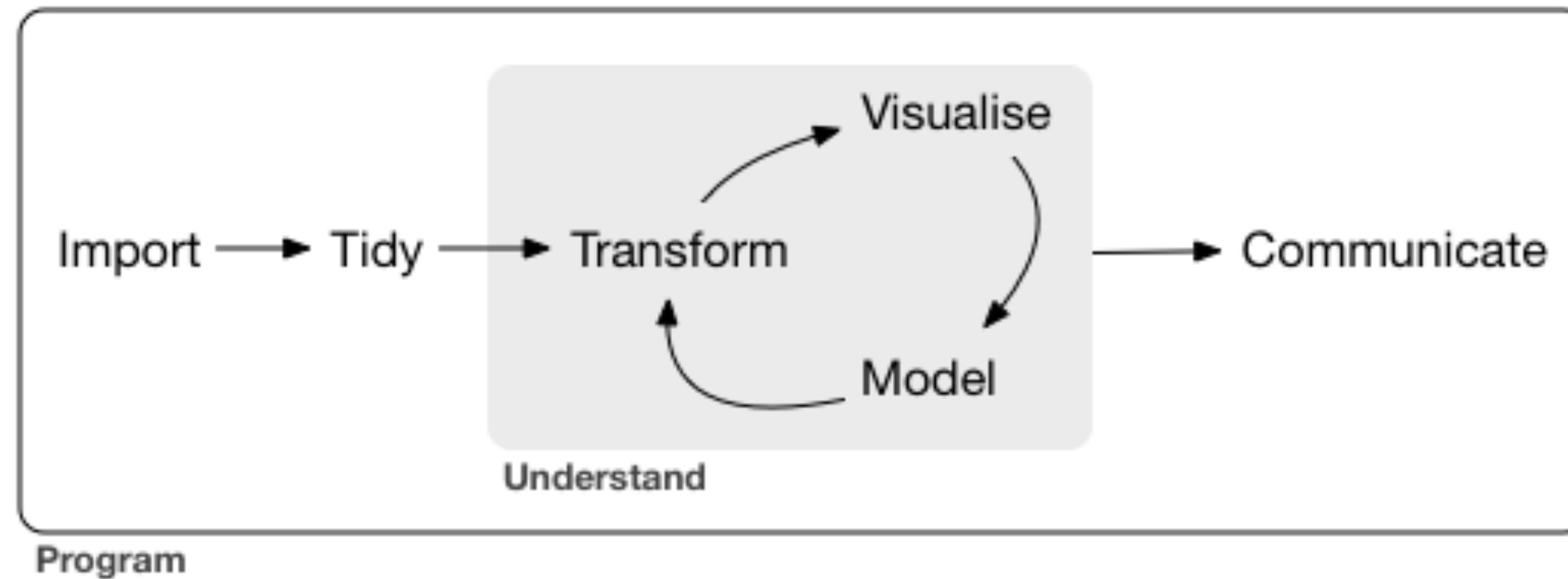
- What's the fuzz about this R language?
 - The philosophy of tidy data:
 - import, manipulate, visualize, communicate
- A grammar of graphics: ggplot2
- Introducing RStudio
- Fancy ggplot2 graphs and some fancy friends:
 - from facets
 - via waterfalls and radar
 - to animated plots and yet some extras
- 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 **Ross Ihaka** and **Robert 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 for Mac OS, Linux, Windows, supported by R Consortium
- Extensible through user-created packages, > 19000 available on CRAN
- Commercial support, e.g. since 2007 by Revolution Analytics, acquired by Microsoft in 2015, now provide Microsoft R Open, R Server
- IDEs: R.App, RStudio, MS R Tools for Visual Studio (< version 2019)
- 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.

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
 - ...

50 ways to show your data

- facets
show small multiples, apply the big picture to subsets of your data
- violins
show more detail of the data distribution than a box plot
- lollipop charts [ggalt]
alternative to charts with lots of nearly-similar-size bars,
on screen: reduce moiré, on paper: reduce waste of ink
- encircling areas [ggalt]
draw attention to certain regions or groups of points,
visualize clustering

50 ways to show your data

- jitter plots / counts charts
show information otherwise hidden by overlapping points
- avoiding overlap of text labels [ggrepel]
avoid overlap of text labels
- parliament diagrams [ggpol]
by seat number or vote distribution
- tree maps [treemapify]
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
- correlograms [(gg)corrplot, corrmorant]
show correlation between several variables at once
- marginal histograms and boxplots [ggExtra]
show distribution details for the dimensions of a scatter plot
- radar charts [ggradar]
compare multiple measures for few items

50 ways to show your data

- maps [ggmap]
include geospatial information
- animated plots [gganimate]
automatically show development over time
- Chernoff faces [ggChernoff]
everyone likes smilies, don't they?
- More extras:
facet zooming [ggforce], diverse themes [ggthemes, ggTech], interactive:
tooltips [ggiraph]

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://posit.co/download/rstudio-desktop/>
- 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.konstanz.de

50 ways to show your data

Thank you for your time and interest & keep in touch:

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

 de.linkedin.com/in/derfredo

 <https://techhub.social/@DerFredo>

This file and the demo scripts can be found at:

<https://bit.ly/DerFredoNDCL23>

