

50 ways to show your data

Presented by Thomas Hütter

for DataSaturday Malta on April 24th, 2021



Thank you to our Sponsor



Malta

Microsoft Data Platform
User Group



Explore

Everything our User
Group Has To Offer



**Get involved in
our Meetup**

<https://bit.ly/3auvRnD>



**Follow our page
on Facebook**

<https://bit.ly/2P9sqLy>



**Join the conversation
on our Facebook group**

<https://bit.ly/2QqAWX4>



**Follow our
Videos on
Youtube**

<https://bit.ly/3n8l5bP>

Thomas Hütter

DataSaturday Malta
2021-04-24

50 ways to show your data



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 SQL events around Europe



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

 de.linkedin.com/in/derfredo

 www.xing.com/profile/Thomas_Huetter

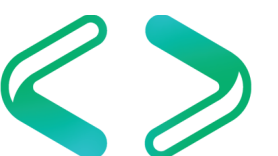


JOIN !
THE DATABASE CONFERENCE



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

Sign up now at sqlkonferenz.de



Agenda

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



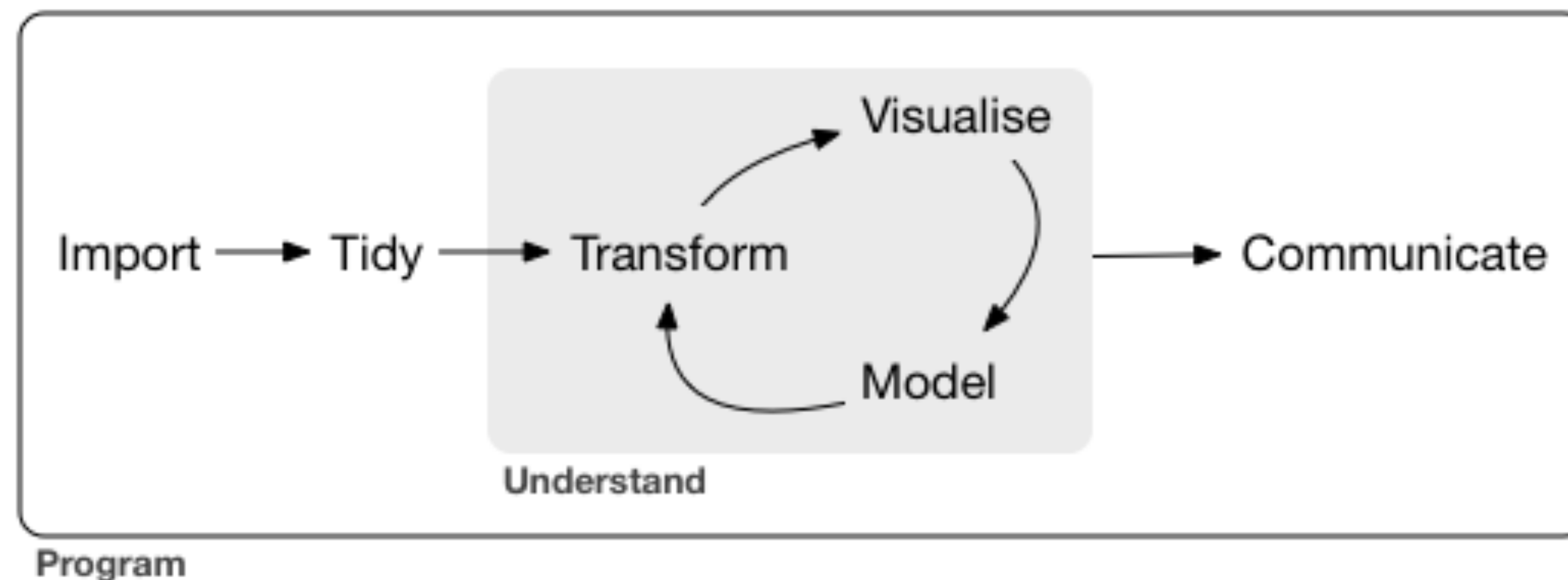
What's the fuzz about this language?

- Programming language for statistical computing, analysis and 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, > 17400 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 in SQL Server, Power BI, Azure ML, Data science VM



The flow of your 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
 - ...



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 bar charts with lots of nearly-similar-sized 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 [treemap]
display hierarchical data as nested rectangles



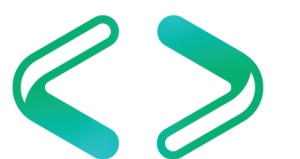
50 ways to show your data

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



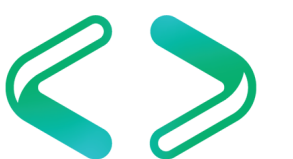
Round-up / conclusions

- With R, a lot is possible in terms of analysis and especially visualizations
- With ggplot2 and its extensions, graphs can be constructed in a very concise manner, according to the grammar of graphics
- Don't overdo/overload your visualizations
- Careful with the number and choice of colours/shapes
- Consider the occasion of your presentation (Chernoff faces are not suitable for every audience)



Credits & resources / inspired by:

- 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>
- Formula 1 data set: www.formel1.de
- Cellphone subscriptions data: data.worldbank.org
- Live parking data: www.kleve.de



50 ways to show your data

Thank you 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/DerFredoMalta21>

