

Visual Forms

S. Santoni¹

¹Bayes Business School

MSc in Business Analytics, 2024/25

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

1 Session #2 Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

Session #3
Wrap Up

3 Session #3 Wrap Up

References

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

1 Session #2 Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

Session #3
Wrap Up

3 Session #3 Wrap Up

References

Grammar of Graphics

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

- The Grammar of Graphics is an analytical framework that considers the design of a chart as a 'bundle of choices'
- Leland Wilkinson proposed GoF in 1999 [4]
- Since then, many data visualization textbooks and packages have built on it
- GoF advantages:
 - GoF facilitates reasoning about a chart design
 - GoF is a discursive tool that help individuals discuss charts

Software Adopting GoF

Python's Plotnine

Visual Forms

S. Santoni

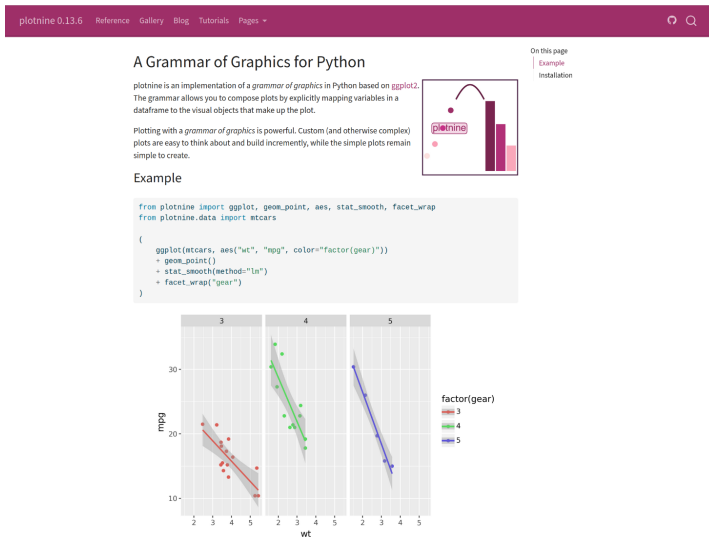
Session #2 Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3 Wrap Up

References



Software Adopting GoF

Python's Vega-Altair

Visual Forms

S. Santoni

Session #2 Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3 Wrap Up

References



Vega-Altair is a declarative visualization library for Python. Its simple, friendly and consistent API, built on top of the powerful [Vega-Lite](#) grammar, empowers you to spend less time writing code and more time exploring your data.

Getting Started

In the Getting Started section you can find installation instructions and a high-level overview of the main concepts.

User Guide

Check out the User Guides for in-depth information on the key concepts of Vega-Altair.

Examples

The Examples gallery contains a selection of different visualizations which you can create with Vega-Altair.

API

The API reference guide contains detailed information on all of Vega-Altair's methods and classes.

Software Adopting GoF

Julia's Gadfly

Visual Forms

S. Santoni


Session #2 Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3 Wrap Up

References



GADFLY
Gadfly.jl

Home

- Package features
- Installation
- Compilation

Tutorial

Manual

- Plotting
- Compositing
- Backends
- Themes

Gallery

- Geometries
- Guides
- Statistics
- Coordinates
- Scales
- Shapes

Library
Version: v1.4.0

Home

Gadfly.jl

Gadfly is a system for plotting and visualization written in [Julia](#). It is based largely on Hadley Wickhams's [ggplot2](#) for R and Leland Wilkinson's book [The Grammar of Graphics](#). It was [Daniel C. Jones'](#) brainchild and is now maintained by the community. Please consider [citing it](#) if you use it in your work.

Package features

- Renders publication quality graphics to SVG, PNG, Postscript, and PDF
- Intuitive and consistent plotting interface
- Works with [Jupyter](#) notebooks via [IJulia](#) out of the box
- Tight integration with [DataFrames.jl](#)
- Interactivity like panning, zooming, toggling powered by [Snap.svg](#)
- Supports a large number of common plot types

Installation

The latest release of Gadfly can be installed from the Julia REPL prompt with

```
julia> ]add Gadfly
```

The closing square bracket switches to the package manager interface and the `add` command installs Gadfly and any missing dependencies. To return to the Julia REPL hit the `delete` key.

From there, the simplest of plots can be rendered to your default internet browser with

```
julia> using Gadfly
julia> plot(y=[1,2,3])
```

Now that you have it installed, check out the [Tutorial](#) for a tour of basic plotting and the various manual pages for more advanced usages.

Compilation

R's ggplot2

S. Santoni

Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

References

ggplot2

3.5.1

Get started


Reference


News

Articles

Extensions

Search for





ggplot2

Overview

Installation

ggplot2 is a system for declaratively creating graphics, based on [The Grammar of Graphics](#). You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

```
# The easiest way to get ggplot2 is to install the whole tidyverse:
install.packages("tidyverse")

# Alternatively, install just ggplot2:
install.packages("ggplot2")

# Or the development version from GitHub:
# install.packages("pak")
pak::pak("tidyverse/ggplot2")
```

LINKS

[View on CRAN](#)

[Browse source code](#)

[Report a bug](#)

[Learn more](#)

[Extensions](#)

LICENSE

[Full license](#)

MIT + file [LICENSE](#)

COMMUNITY

[Contributing guide](#)

[Code of conduct](#)

CITATION

[Citing ggplot2](#)

DEVELOPERS

[Hadley Wickham](#)

Author

[Winston Chang](#)

Author

Porting GoF to Specific Software

Mapping between Wilkinson's GoF and ggplot2's GoF

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

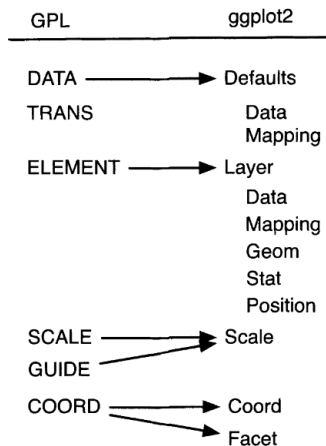
Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References



Source is [2]

ggplot2's Internal Structure

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

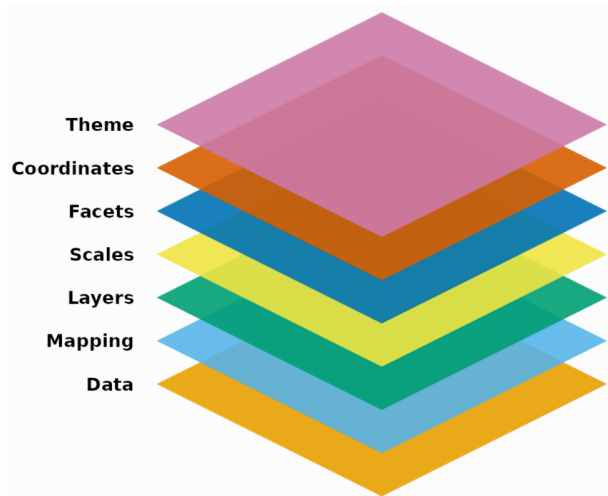
Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References



Source is <https://ggplot2.tidyverse.org/articles/ggplot2.html>

A Minimal ggplot2 Snippet

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

```
1 library(ggplot2)
2
3 ggplot(
4   data = TIDYDATA,
5   mapping = aes(x = COL_1, y = COL_2, colour = COL_3)
6 ) +
7   geom_point()
8
```

Data

Visual Forms

S. Santoni

Session #2 Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3 Wrap Up

References

“The system works best if the data is provided in a tidy format, which briefly means a rectangular data frame structure where rows are observations and columns are variables.”

“As the first step in many plots, you would pass the data to the `ggplot()` function, which stores the data to be used later by other parts of the plotting system.”

```
1  ggplot(data = TIDYDATA)
2
```

Mapping

Visual Forms

S. Santoni

Session #2 Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3 Wrap Up

References

“The mapping of a plot is a set of instructions on how parts of the data are mapped onto aesthetic attributes of geometric objects. It is the ‘dictionary’ to translate tidy data to the graphics system.”

```
1 ggplot(data = TIDYDATA, mapping = aes(x = COL_1, y = COL_2))  
2
```

Layers

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

“The heart of any graphic is the layers. They take the mapped data and display it in something humans can understand as a representation of the data. Every layer consists of three important parts:

- 1 The geometry that determines how data are displayed, such as points, lines, or rectangles*
- 2 The statistical transformation that may compute new variables from the data and affect what of the data is displayed*
- 3 The position adjustment that primarily determines where a piece of data is being displayed.”*

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

3 Session #3 Wrap Up

What Are the Core Visual Forms in Data Visualization?

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3
Wrap Up

References



A Taxonomy of Visual Forms

Visual Forms

S. Santoni

Session #2 Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

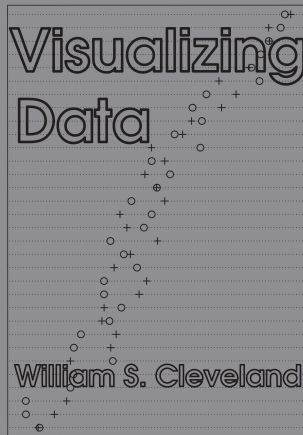
Hypervariate plots

Session #3 Wrap Up

References

Cleveland [1] proposes a taxonomy of the visual forms based on the cardinality of a plot's **mapping** Φ :

- $||\Phi|| = 1 \rightarrow$ univariate plots
- $||\Phi|| = 2 \rightarrow$ bivariate plots
- $||\Phi|| = 3 \rightarrow$ trivariate plots
- $||\Phi|| \geq 4 \rightarrow$ hypervariate plots



Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

3 Session #3 Wrap Up

A Toy Dataset

'Singer' – singer's height and vocal extension data (N=235)

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

Height	Voice part
64	Soprano 1
62	Soprano 2
65	Alto 2
67	Alto 1
72	Tenor 1
69	Tenor 2
75	Basso 1
74	Basso 2
...	...

Quantile Plot

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

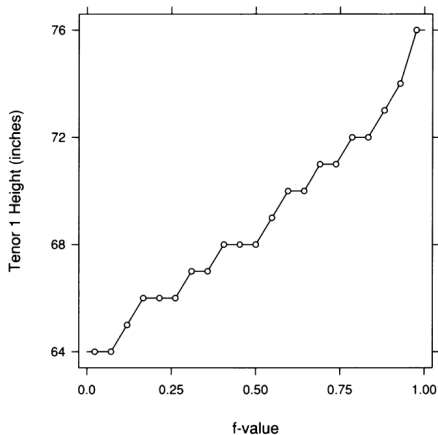
Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References



- Quantile plots are essential to visualize a distribution, i.e., the collection of positions in the data
- The quantile f , $q(f)$, of a set of data is a value along the measurement scale of the data with the property that approximately a fraction of the data are less than or equal to $q(f)$
- The property has to be approximate because there might not be a value with exactly a fraction f of the data less than or equal to it

Quantile-Quantile Plot

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

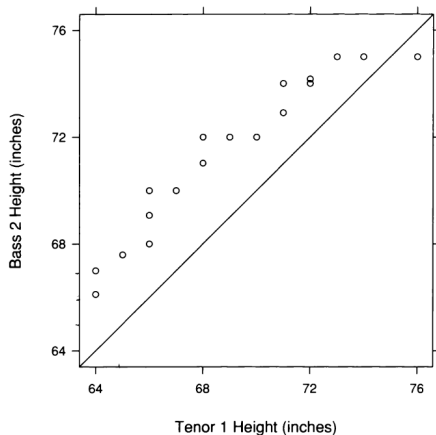
Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References



- The quantile-quantile plot, or q-q plot [3], is a powerful visualization method for comparing the distributions of two or more sets of univariate measurements
- When distributions are compared, the goal is to understand how the distributions shift in going from one data set to the next
- For the singers, the goal is to understand how the height distributions shift with voice part

A QQ Plot Extension: The Pairwise QQ Plot

Visual Forms

S. Santoni

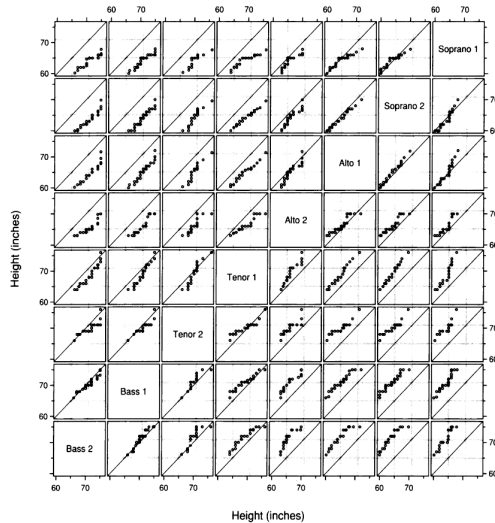
Session #2
Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3
Wrap Up

References



The Boxplot

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

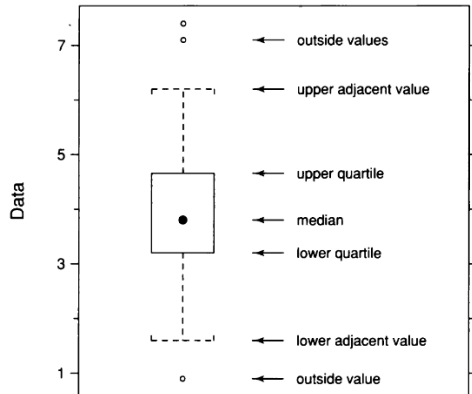
Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References



Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- **Bivariate plots**
- Trivariate plots
- Hypervariate plots

3 Session #3 Wrap Up

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- Bivariate plots
- **Trivariate plots**
- Hypervariate plots

3 Session #3 Wrap Up

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

3 Session #3 Wrap Up

Outline

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots

Bivariate plots

Trivariate plots

Hypervariate plots

Session #3
Wrap Up

References

1 Session #2 Wrap Up

2 Visual Forms

- Univariate plots
- Bivariate plots
- Trivariate plots
- Hypervariate plots

3 Session #3 Wrap Up

Visual Forms

S. Santoni

Session #2
Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3
Wrap Up

References

Time to wrap up!

References

Visual Forms

S. Santoni

Session #2 Wrap Up

Visual Forms

Univariate plots
Bivariate plots
Trivariate plots
Hypervariate plots

Session #3 Wrap Up

References

- [1] William S. Cleveland. *Visualizing Data*. Murray Hill, N.J. : [Summit, N.J.clevelandVisualizingData1993: At&T Bell Laboratories ; Published by Hobart Press, 1993. 360 pp.
- [2] Hadley Wickham. “A Layered Grammar of Graphics”. In: *Journal of Computational and Graphical Statistics* 19.1 (Jan. 2010), pp. 3–28.
- [3] Martin B Wilk and Ram Gnanadesikan. “Probability plotting methods for the analysis of data”. In: *Biometrika* 55.1 (1968), pp. 1–17.
- [4] Leland Wilkinson. *The Grammar of Graphics*. Springer, 2012.