

Have a clear  
purpose

Show the  
data  
clearly



Make the  
message  
obvious

# Effective Visual Communication for Quantitative Scientists

Mark Baillie

September 10<sup>th</sup>, 2021

<https://graphicsprinciples.github.io/>

# STRATOS Visualization panel

“Visualization and the use of graphics can help at every stage of an analysis, from the planning and design of an experiment, the very first data explorations, through to the communication of conclusions and recommendations. Visualization is more than "plotting data"; it can lead to a deeper understanding and inform next steps.

The role of the STRATOS visualization panel is to promote the use of good graphical principles for effective visual communication, providing guidance and recommendations covering all aspects from the design, implementation and review of statistical graphics.”

<http://www.stratos-initiative.org>

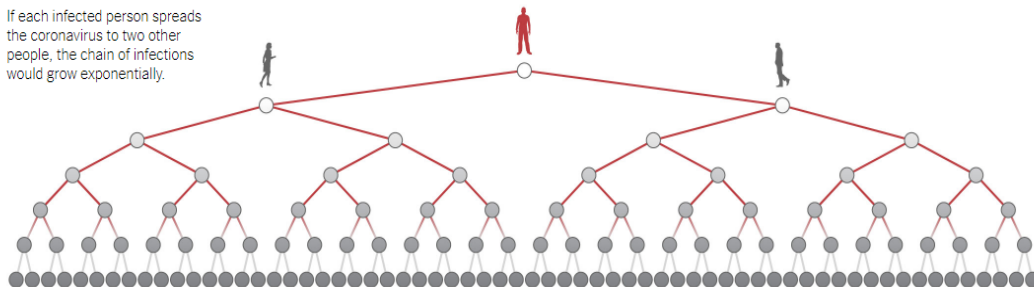


# Effective visualisation is important

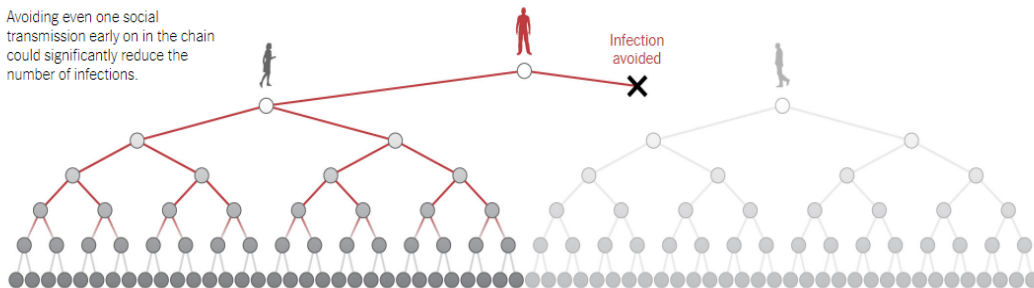
## Cutting a Link in the Chain of Transmission

A simple tree diagram shows how limiting contacts early might prevent many infections.

If each infected person spreads the coronavirus to two other people, the chain of infections would grow exponentially.



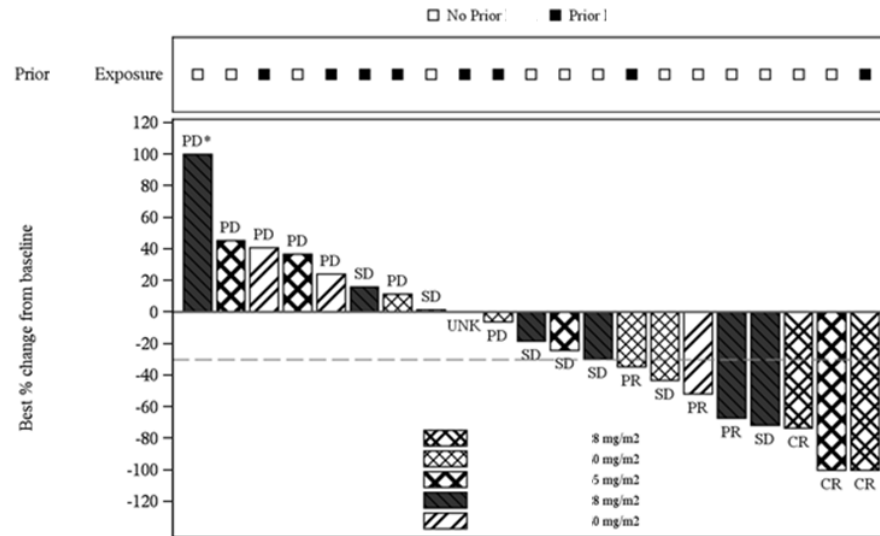
Avoiding even one social transmission early on in the chain could significantly reduce the number of infections.



By Jonathan Corum

# We are not always good at it

Figure 11-1 (Page 1 of 1)  
 Best percentage change from baseline in sum of longest diameters and best overall response as per investigator by prior treatment (Full analysis set)

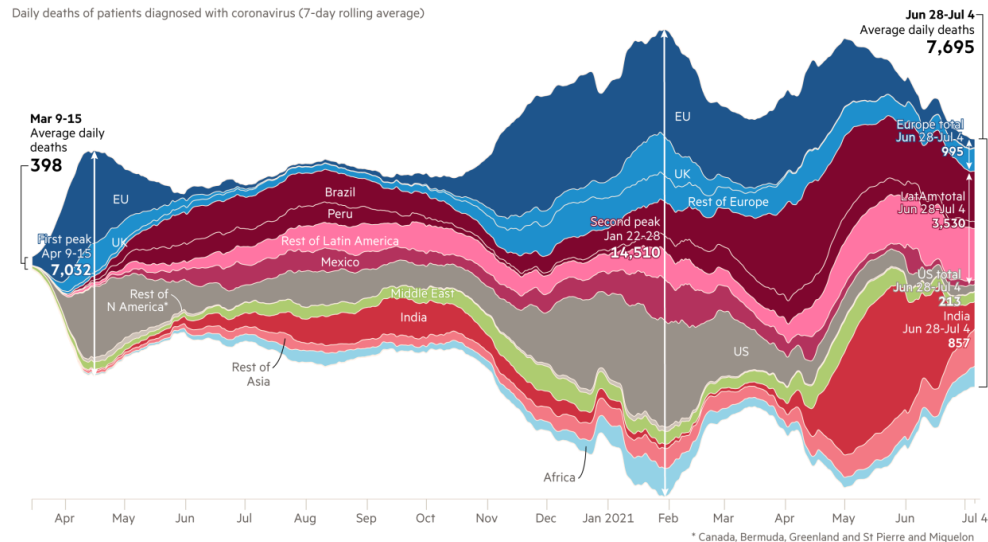


-. \* Denotes the percentage change from baseline greater than 100.  
 Source: Table 11-4, Listing 14.2-1.2 and Listing 16.2.4-1.5

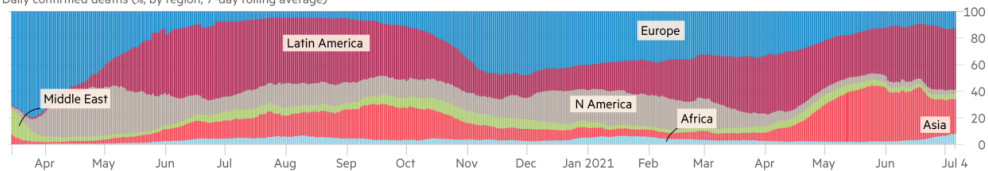
# Beautiful but effective?

The global third wave of Covid deaths is easing

Daily deaths of patients diagnosed with coronavirus (7-day rolling average)



Daily confirmed deaths (% by region, 7-day rolling average)



FT graphic: Steven Bernard / @sdbernard

Sources: FT analysis of data from Johns Hopkins CSSE, WHO, UK government coronavirus dashboard, Swedish Public Health Agency

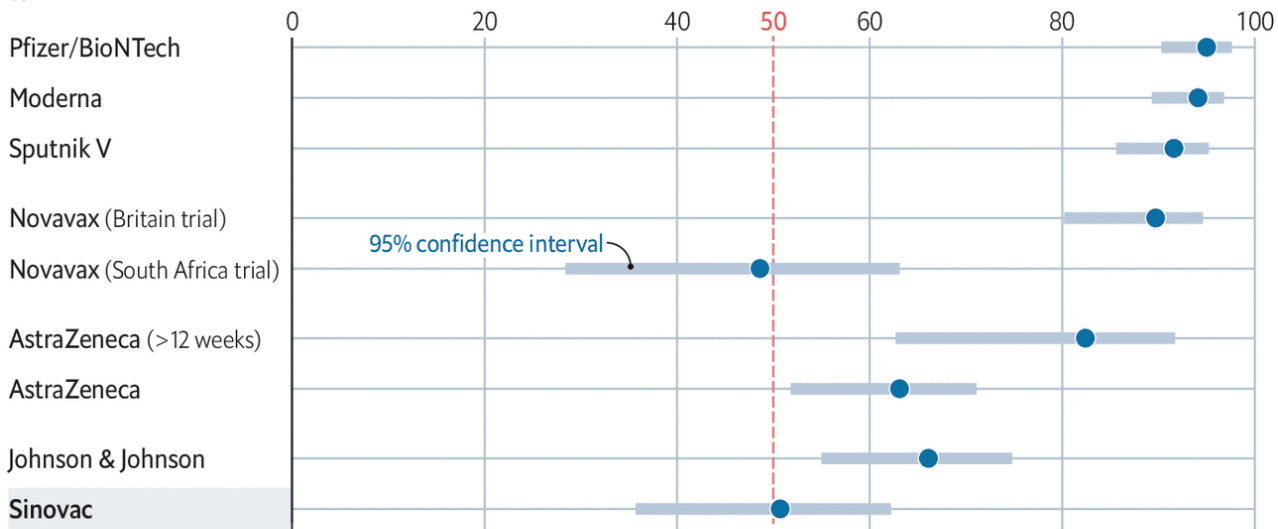
© FT

# Beautiful and effective?

## Making the cut

Covid-19 vaccine efficacy results in phase-three trials\*

%



Source: Airfinity

\*Only trials with known confidence intervals

The Economist

# Principles for effective visual communication

## Graphical Principles Cheat Sheet

Authors: Mark Bailille,<sup>1</sup> Alison Margoskee,<sup>2</sup> Baldur Magnusson,<sup>1</sup> Andrew Wright,<sup>1</sup> Ruqan You,<sup>2</sup> Ivan-Toma Vranesic,<sup>1</sup> Marc Vandemeulebroeck<sup>1</sup>  
 Affiliations: <sup>1</sup>Novartis Pharma AG, Basel, Switzerland; <sup>2</sup>Novartis Institutes for Biomedical Research, Cambridge, MA, United States; <sup>3</sup>Novartis Institutes for Biomedical Research, Shanghai, China

### Communication

Effective visualizations communicate complex statistical and quantitative information facilitating insight, understanding, and decision-making.

But what is an effective graph?

This cheat sheet provides general guidance and points to consider.

### Principles of Effective Graphic Design

**Proximity** – group related elements together

**Alignment** – elements on the same vertical or horizontal plane are perceived as having similar properties

**Simplicity** – cut anything superfluous, only include elements that add value, limit to 2-3 colors or fonts

**White space** (empty space) – use white space to minimize distraction & provide clarity

**Legibility** – sans serif fonts are easier to read, use color for emphasis instead of a new typeface

**Color** – select colors that present enough contrast to make the graph legible. Choose monochromatic color schemes to prevent clashing. Use dark colors and accent colors to emphasize important information

**Visual Hierarchy** – use color, font, image size, typeface, alignment & placements create a viewing order

**Focal Points** – primary area of interest that immediately attracts the eye, emphasize the most important concept and make it your focal point

**Repetition** – repeating elements can be visually appealing, repeated shapes, labels, colors

**Familiarity** – using familiar styles, icons, navigation structure makes viewers feel confident

**Consistency** – be consistent with heading sizes, font choices, color schemes, and spacing. Use images with similar styles

### Selecting the right base graph

Consider if a standard graph can be used by identifying suitable designs based on the:

- (1) **purpose** (i.e. message to be conveyed or question to answer) and
- (2) **data** (i.e. variables to display).

Example plots categorized by purpose:

Revision	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Marginalize
Box plot Baseline	Scatter plot	Horizontal bar chart	Boxplot	Kaplan Meier	Stacked bar chart	Vertical bar chart
Waterfall	Heat map	Dendrogram	Histogram	Line plot	Tree map	Forest plot

Utilize the **Graph Gallery** ([gobargallery](#)). There are many effective base designs that can be adapted to your purpose.

### Implementation Considerations

Plot cause on the x-axis and effect on the y-axis. Use the **standard** orientation in order to avoid misinterpretation.

Aspect ratio can influence interpretation. Aim for a 4:5 degree angle of change to avoid over-interpretation of slope.

Use position for comparisons rather than length (i.e. dots instead of bars), especially for non-linear scales (e.g. log scales or % change).

Do not type too small or too condensed. Break long titles into two lines. Shift or adjust size of labels that overlap.

Do not plot log-normally distributed variables on a linear scale (e.g. hazard ratio, AUC, CL).

When displaying data measured on the same scale, also plot them on the same scale for easy comparison.

Connected data imply continuity. Do not connect data across a disconnected or uneven time scale.

Visits displayed close together are perceived to be closer in time. Space the visits proportional to the time between them in order to avoid confusion (linear, linear or piecewise).

Plot data and inferences to support stories about the models.

### Legibility and Clarity

Effective graphs stand alone. They use titles, annotations, labels, shapes, colors, and textures to deliver important information.

Label axes with clear measurement units and provide annotations that support the message.

Use font size to create hierarchy (e.g. set titles 3pt larger than all other labels to make them more prominent).

Keep the font style simple, sans serif is easier to read.

Display text with enough contrast to be visible. Favor the use of dark on light instead of light on dark whenever possible.

Bold or italics should only be used for layering or emphasis. Emphasizing everything means nothing gets emphasized.

Try not to text at an angle, this decreases readability. Think of alternative solutions such as transposing the graph.

### Good graph checklist

Clear Communication	Implementation Considerations
<ul style="list-style-type: none"> <li>Is the message of the graph as clear as possible?</li> <li>Is it easy for someone unfamiliar with the data to interpret the graph?</li> <li>Are the patterns/relationships easily identifiable?</li> <li>Is the graph tailored to its primary purpose and audience?</li> <li>Is the correct graph type used?</li> </ul>	<ul style="list-style-type: none"> <li>Are multiple panels plotted on the same scale?</li> <li>Are logarithmically distributed variables plotted on a log scale?</li> <li>Are common baselines used wherever identifiable?</li> <li>Does the orientation of the axes aid interpretation?</li> <li>Does the aspect ratio allow the reader to see variations in the data?</li> <li>Are data across a disconnected time scale kept disconnected?</li> <li>Are data spaced proportionally to the actual time interval (instead of according to visit number)?</li> <li>Are data and inferences plotted to support stories about models?</li> <li>Are numbers of patients by group reported if this adds context?</li> </ul>
Facilitating Comparisons	Legibility and Clarity
<ul style="list-style-type: none"> <li>Are elements to be compared grouped together?</li> <li>Are labels placed next to data instead of in legends?</li> <li>Have categories been ordered for easy comparison?</li> <li>Can the plot be read without doing mental calculations?</li> <li>Are the estimates of interest plotted (e.g. mean differences with confidence intervals)?</li> </ul>	<ul style="list-style-type: none"> <li>Can all graphical elements be seen?</li> <li>Does the graph have a clear title, axis labels, annotations and data units?</li> <li>Can the font be read without eye strain or effort?</li> <li>Are sans-serif fonts used?</li> <li>Do text sizes have correct hierarchy (big to small, main text to subtext)?</li> <li>Are the elements of the graph clearly labeled (e.g. p, q, points, error bars, shaded regions)?</li> <li>Are labels oriented horizontally where possible?</li> </ul>
Facilitating Comparisons	Color for emphasis or distinction
<ul style="list-style-type: none"> <li>Are elements to be compared grouped together?</li> <li>Are labels placed next to data instead of in legends?</li> <li>Have categories been ordered for easy comparison?</li> <li>Can the plot be read without doing mental calculations?</li> <li>Are the estimates of interest plotted (e.g. mean differences with confidence intervals)?</li> </ul>	<ul style="list-style-type: none"> <li>Are graphical elements displayed in a dark color on a light background?</li> <li>Are light lines drawn with a thin line and a light color such as grey?</li> <li>Are colors used sparingly (e.g. max 3)?</li> <li>Do all elements in the graph have a purpose (e.g. colors, textures, grid lines)?</li> <li>Are the same colors used to mean the same thing in a series of graphs?</li> </ul>

### Effectiveness Ranking

A graph is a representation of data that visually encodes numerical values into attributes such as lines, symbols and colors. The Cleveland-AKCUI scales can be used to select the most effective attribute for your purpose.

Volume	Color hue	Depth	3d	Color intensity	Area	Position on unaligned scale	Position on common scale
Least accurate							Most accurate

volume charts, poorly designed pie charts, multivariate density plots, heat maps, bubble charts, line graphs, bar charts, parallel waterfall, multiple pie charts, stacked bar charts, small multiple plots, coordinate charts, dot plots, bar charts, parallel multiple plots

### Facilitating Comparisons

**Proximity improves association**

Place labels next to data instead of using legends

Group together elements to be compared directly

**Ease visual inspection**

Order values to help compare across many categories

Judgments are easier to make on a common vertical scale

**Reduce mental arithmetic**

Plot the final comparison e.g. mean difference not two means

Use reference lines and other visual anchors

### Color for emphasis or distinction

Restrained use of color is highly effective in organizing a narrative and calling attention to certain elements.

Think carefully about introducing additional color. Do you really need it?

Do not use color to differentiate between categories of the same variable

Use colors or shades to represent meaningful differences such as positive/negative values, treatments or doses

Be consistent, use the same color to mean the same thing in a series of graphs (e.g. treatment, dose)

Use a bold, saturated or contrasting color to emphasize important details

Emphasize the data by minimizing unnecessary ink, e.g. softer gridlines with a light color

Utilize existing resources for selection of appropriate palettes such as Color brewer or Munsell

### Putting it all together – Remove the clutter & emphasize the message

Creating a graph is an iterative process: produce, review and refine.

Colors, backgrounds, and textures can be removed and gridlines reduced.

It is easier to see differences in position over a difference in length (i.e. a dot over a bar).

Using too many colors can be distracting. Use white background and try using other methods to distinguish different curves.

One solution could be repeating the data in different panels, highlighting individual curves in a darker color.

### Resources

Books:

- E. R. Tuft, The visual display of quantitative information, Connecticut, Graphics Press, 2001
- Chavez, W. S. and Miller, Robert. Digital image generation theory, experimentation and application to the development of graphical methods, JSA, Vol. 79, No. 37, p. 531-554, 1984.
- Lee, Peter. The Elements of Design: Tables and Graphs in English, Clarendon, Oxford, 1999.
- D. M. Wong, The Wall Street Journal Guide to Information Graphics: The Data and Darts of Presenting Data, Facts, and Figures, N. B. Books, Creating More Effective Graphs, Cheat House
- J. Doornik, Trees, maps, and thorems: Effective communication for rational minds. PRINCIPAE.
- N. B. Books, Creating More Effective Graphs, Cheat House

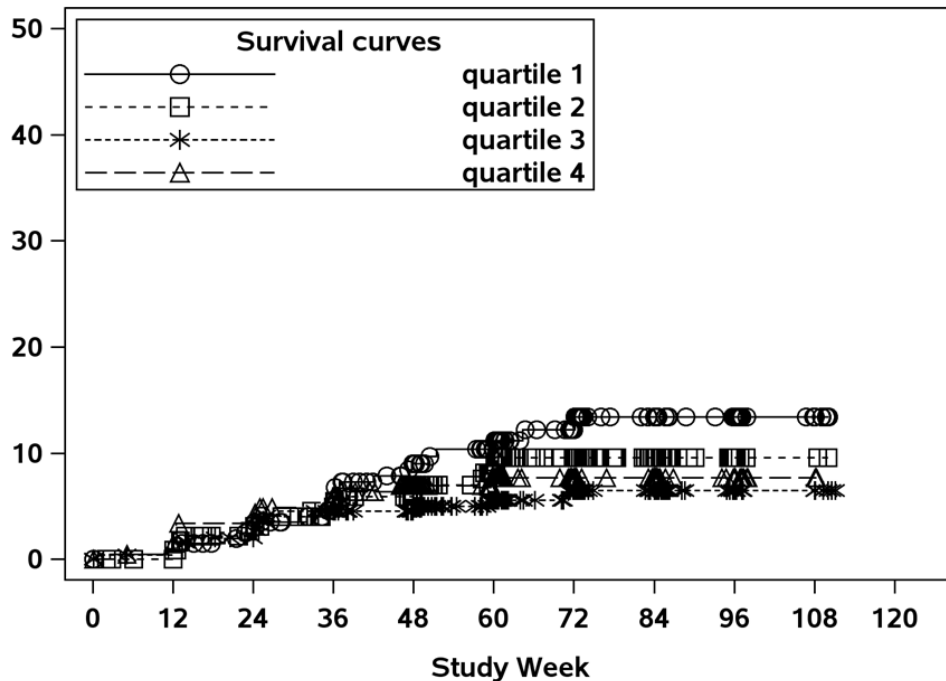
Online resources:

- <http://www.data-visualization.com/> (S. Few)
- <http://www.infographic.com/> (L. Oromoti)
- <http://www.adaptiveresponse.com/> (E. Tuft)
- <http://www.theinformationage.com/> (A. Carro)
- <http://www.creativecommons.org/licenses/by/4.0/> (M. Rutledge)

# This is a continual process

Planned Treatment:

mg





# Three principles for improving visual communication

## Have a clear purpose

- Know the purpose of creating the graph
- Identify the quantitative evidence to support the purpose
- Identify the audience and focus the design to support their needs

## Show the data clearly

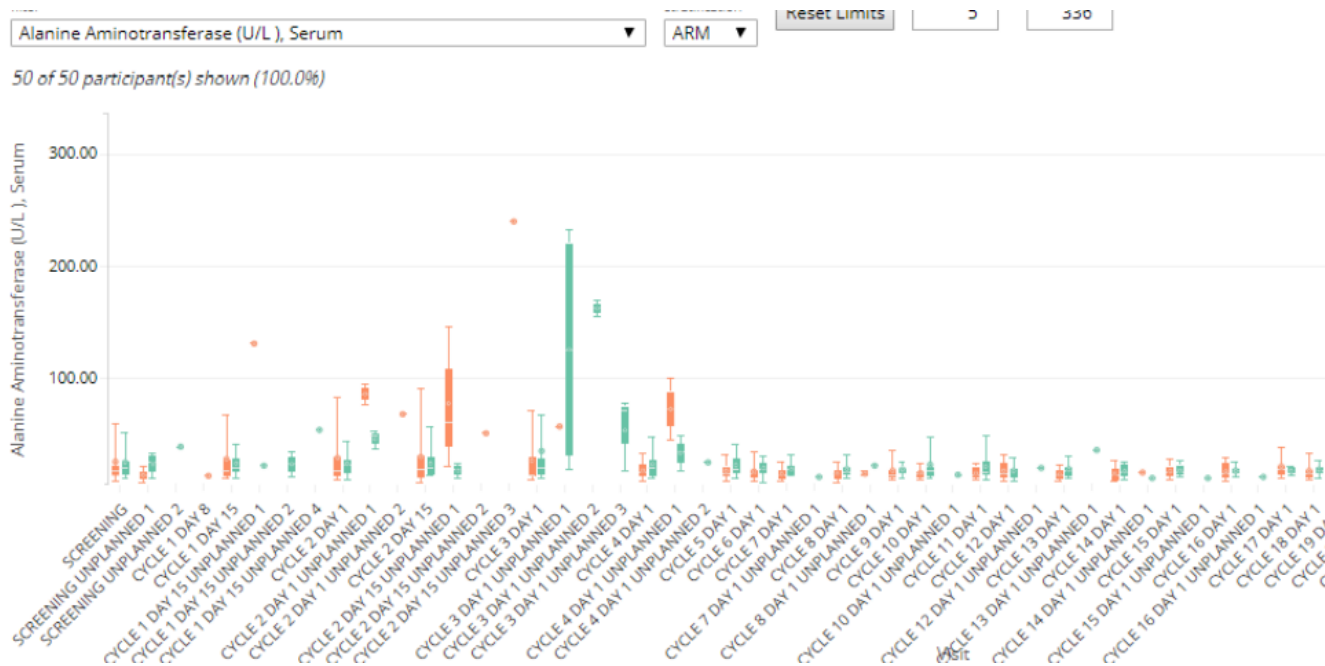
- Choose the appropriate graph type to display your data
- Avoid misrepresentation (use appropriate scales)
- Maximize data to ink ratio (reduce distraction, less is more)

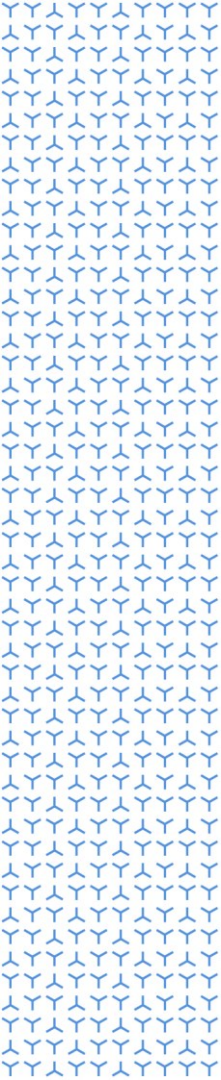
## Make the message obvious

- Use proximity and alignment to aid in comparisons
- Minimize mental arithmetic (e.g. plot the difference)
- Use colors and annotations to highlight important details



# This is a continual process...





Forget about the graph, think about the purpose.



# Law 1

## Have a clear **purpose**

# The 4 areas for a clear purpose

## What is the purpose of the visualization?

- What is the main objective of the visualization?
- List the (scientific) question(s) the visualization is trying to answer. Try to be specific.
- What is the key evidence that is available to answer the question?

## Who is your audience?

- List the primary groups or individuals you will be communicating to.
- If you had to narrow that to a single person, who would that be?
- What does your audience care about?
- What action does your audience need to take?

## What is the importance of this project?

- What are the benefits if your audience acts in the way that you want them to?
- What are the risks if they do not?

## What is the key message (the so what?)

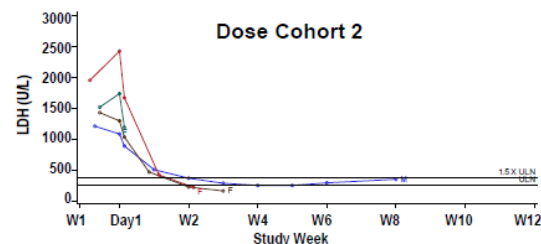
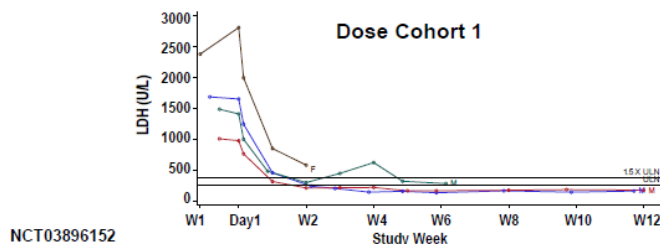
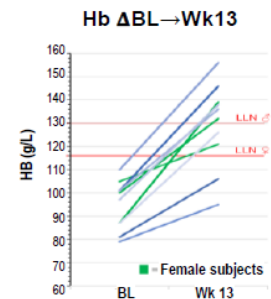
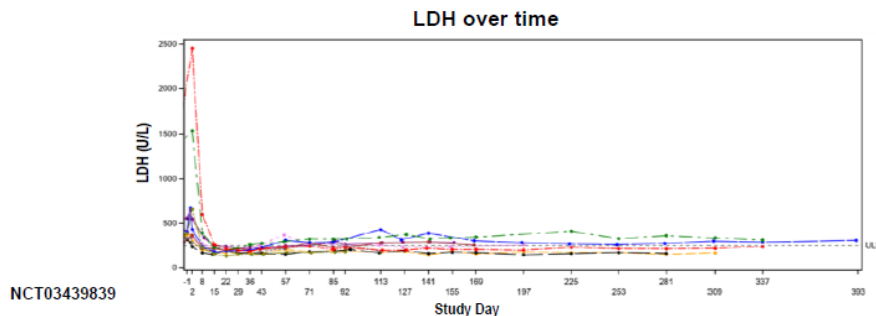
- Write out in a single sentence the key message



# Early Phase 2 data support advancing LNP023 as a front-line treatment for PNH

In a Phase 2 PNH trial, LNP023 add-on to eculizumab in patients with hemolysis delivered consistent LDH normalization and transfusion-free hemoglobin increase in all patients

The ongoing LNP023 monotherapy trial in eculizumab-naïve PNH patients shows early efficacy (LDH↓)



# Purpose



## What is the purpose of the visualization?

What is the main objective of the visualization?

*The visualization is to display supporting evidence that LNP023 has demonstrated proof of concept and is a good candidate to take into phase 3 development.*

List the (scientific) question(s) the visualization is trying to answer. Try to be specific.

- *Is there a decrease in LDH to “normal levels” post LNP023 dose as a mono and combo therapy?*
- *Does LNP023 increase hemoglobin levels?*

What is the key evidence that is available to answer the question?

*Two studies.*

*Two different dose cohorts in one study. Mono and combo.*

*LDH is a surrogate measure of efficacy for PNH.*

*Consistency across gender for Hemoglobin improvement.*

# What type of graph do I want to create?

## EXPLORATORY

“I want to dig into the data”

“I want to get familiar with the data”

“I want to find the story in my data”

The audience is:  
**YOU**

## EXPLANATORY

“I want to communicate the results”

“I want to tell the story behind the data”

The audience is:  
**SOMEONE ELSE**

# Do you want your audience to play 'Where's Wally?'



Credit Andrew Wright, Novartis



**Law 2**  
**Show the data clearly**



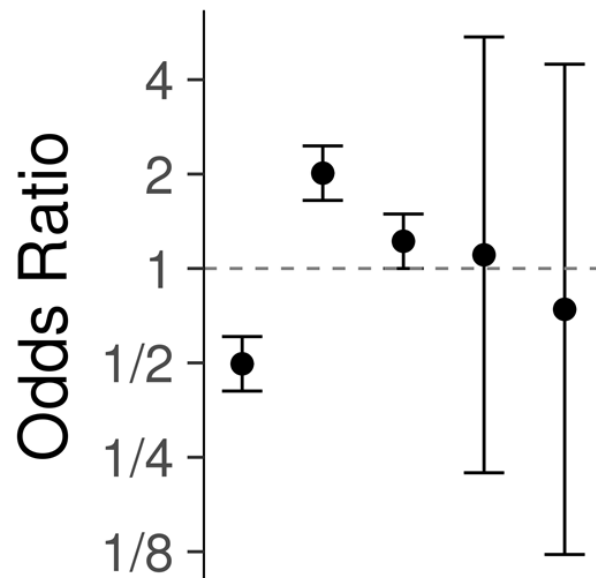
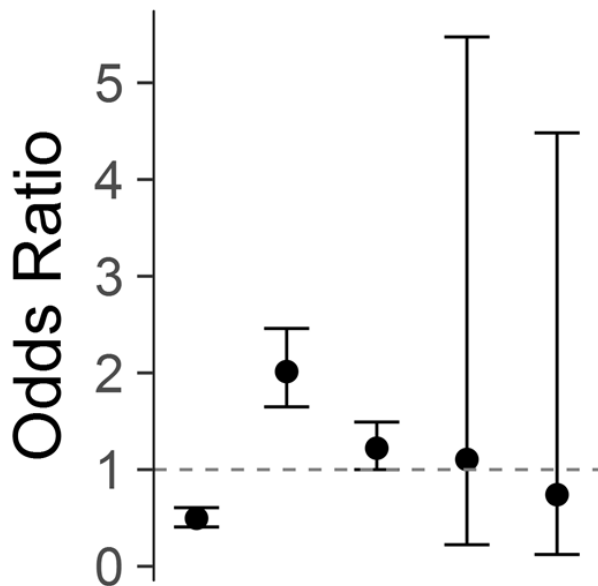
# Show the data clearly



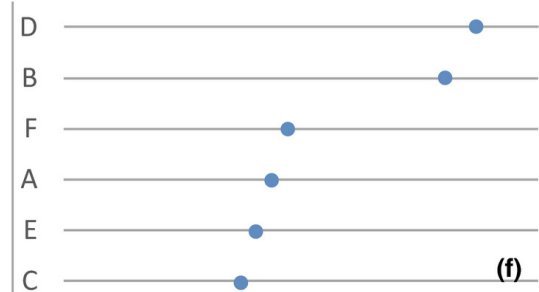
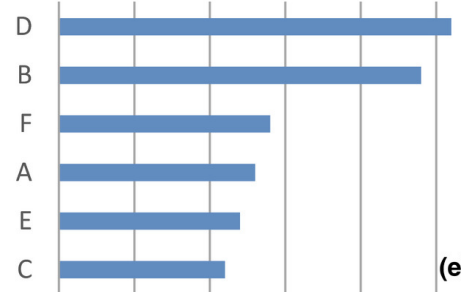
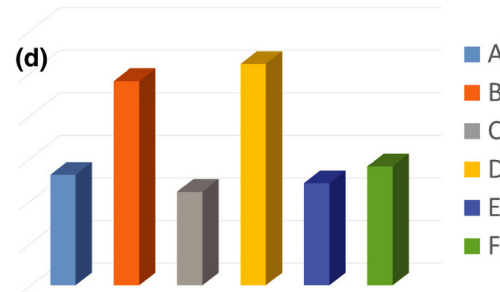
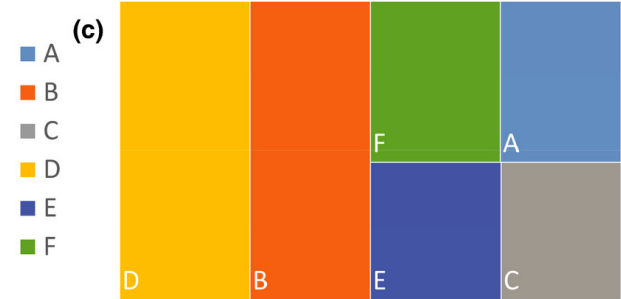
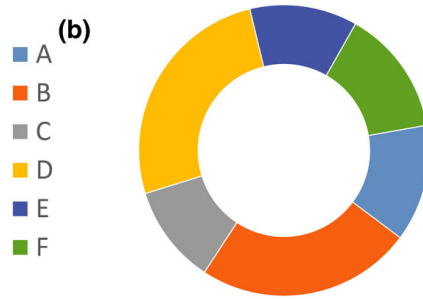
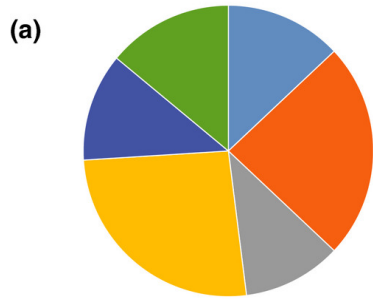
The Guardian

# Choose the right scale for your data

Avoid plotting log-normally distributed variables on a linear scale  
(e.g. hazard ratio, AUC, CL)

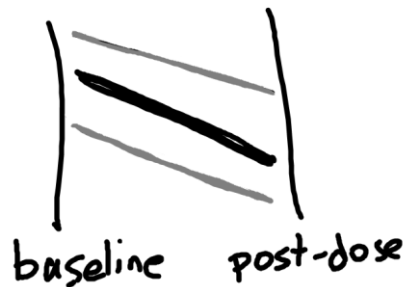
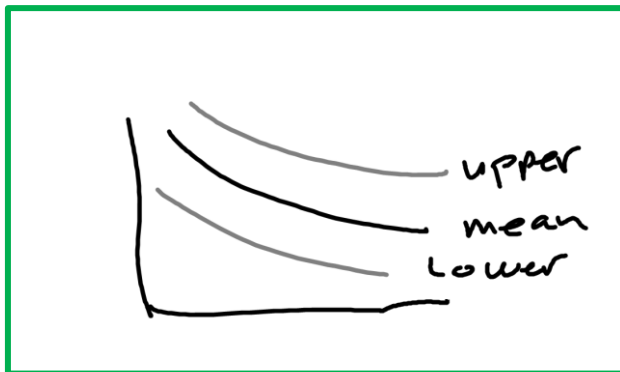


# Choosing the Correct Graph Type Aids in Interpretation



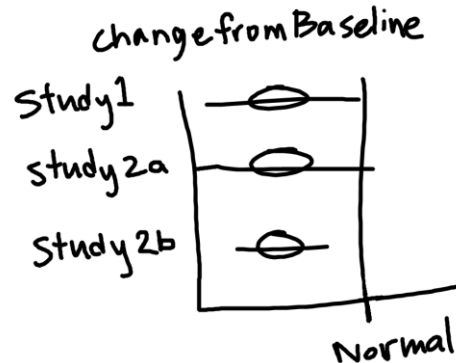
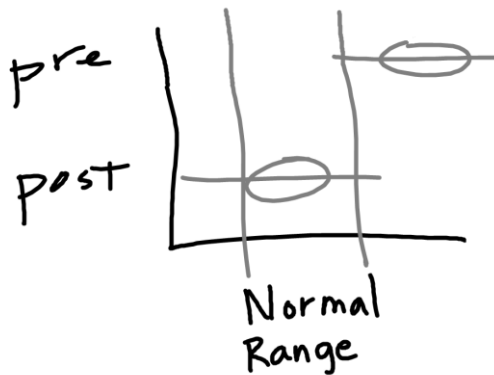
# 2<sup>nd</sup> principle - select the appropriate graph

- Come up with several different ways to display the same information
- Display the key evidence in a way that supports the purpose

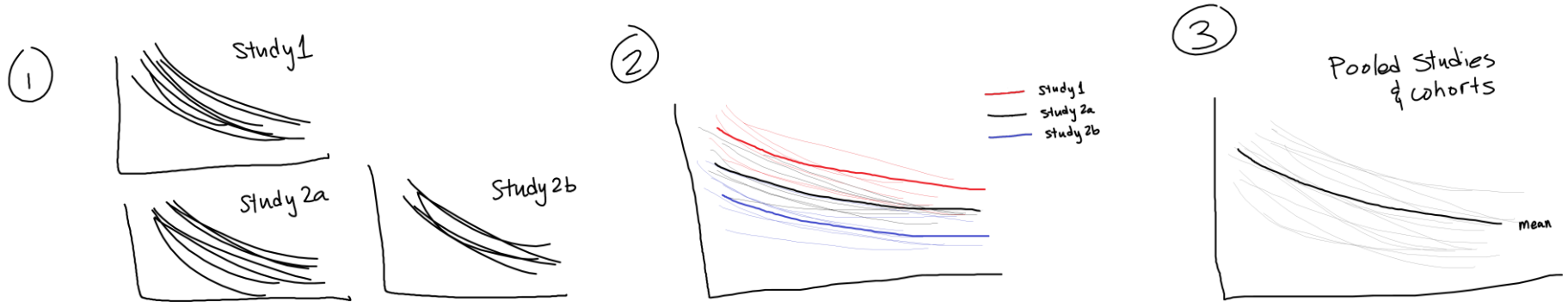


Continuing with the LNP example

- What is the key message: LNP023 reduces LDH levels to normal
- What is the key evidence to support this: Two studies, different dose cohorts, LDH as a surrogate for efficacy



# 2<sup>nd</sup> principle – Iterate and eliminate clutter

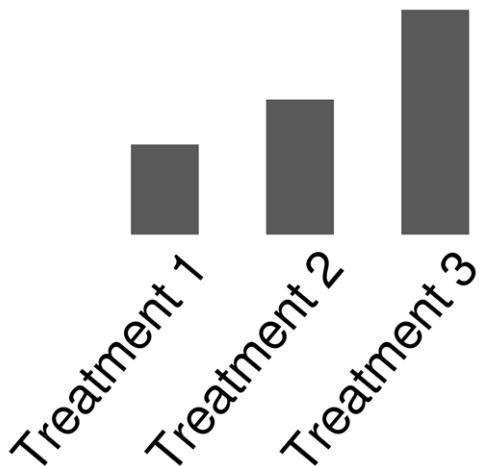


**Law 3**  
**Make the *message* obvious**



# Try not to set text at an angle

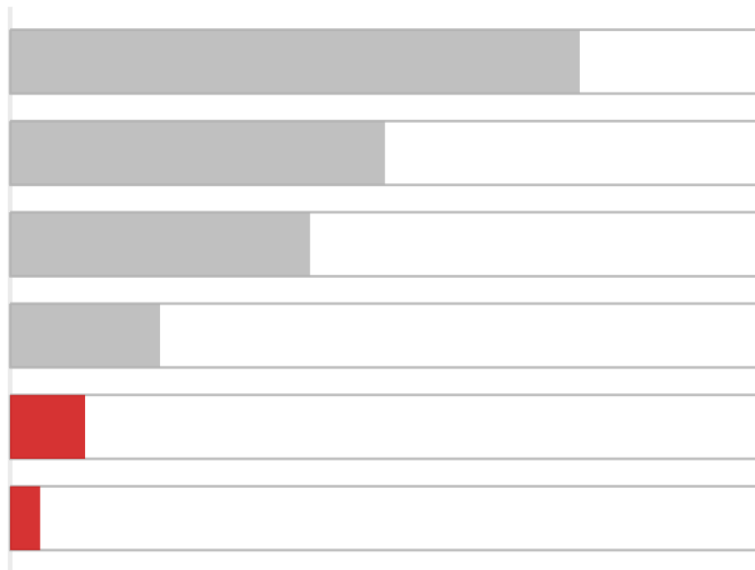
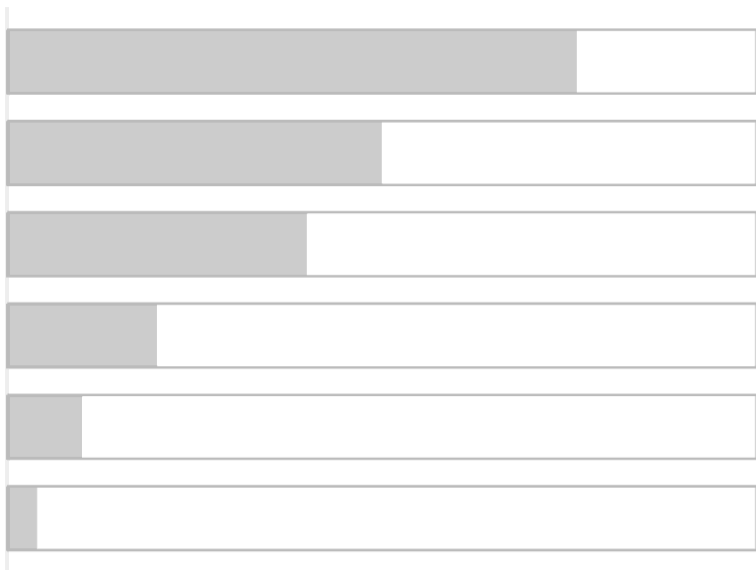
Think of alternatives such as transposing the graph



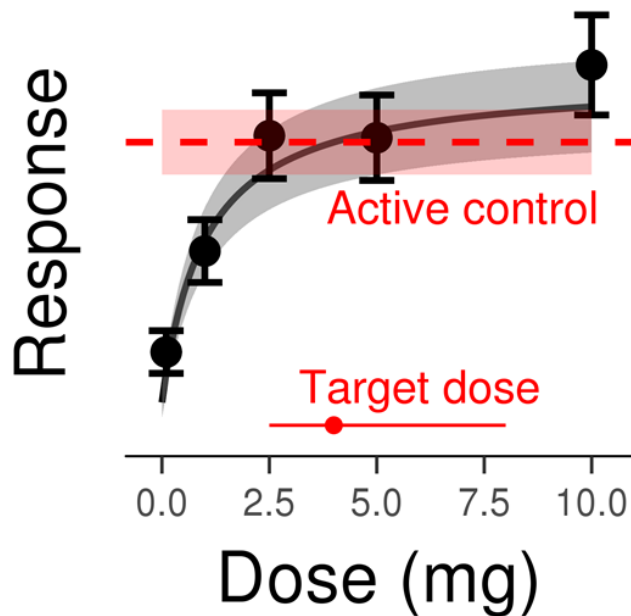
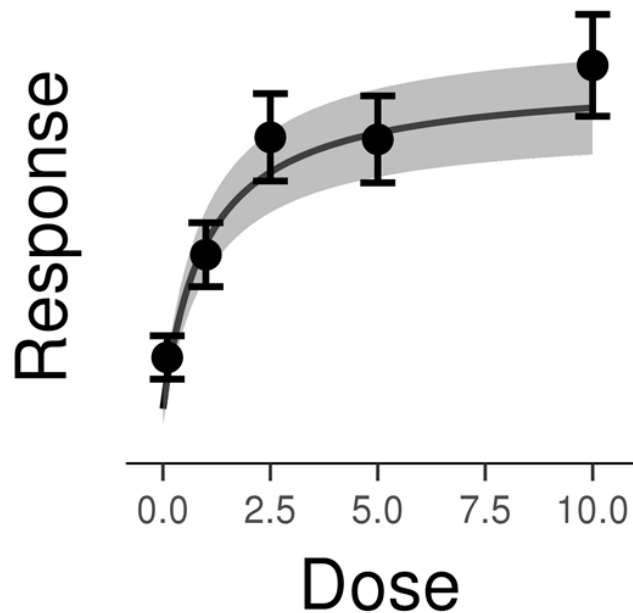


# Only use color when it adds value

Use a bold, saturated or contrasting color to emphasize important details

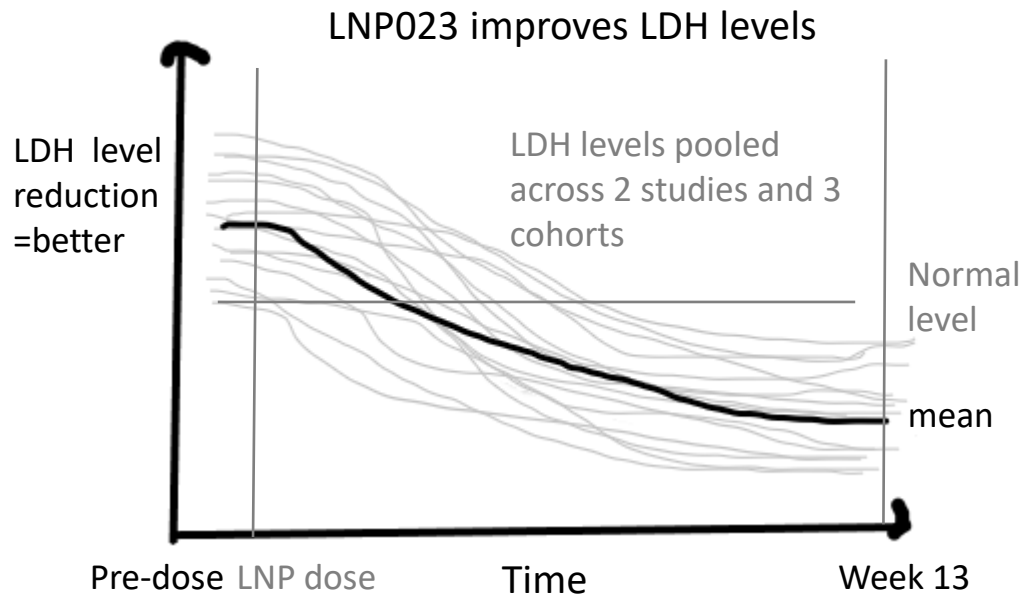


# Use informative labels and annotations to support the message



# 3<sup>rd</sup> principle – draw attention

- Draw the viewer's attention to points of interest
- Use arrows, labels, reference lines to drive home the message
- Make sure to have clear axis labels and informative titles



# Where to find to out more?

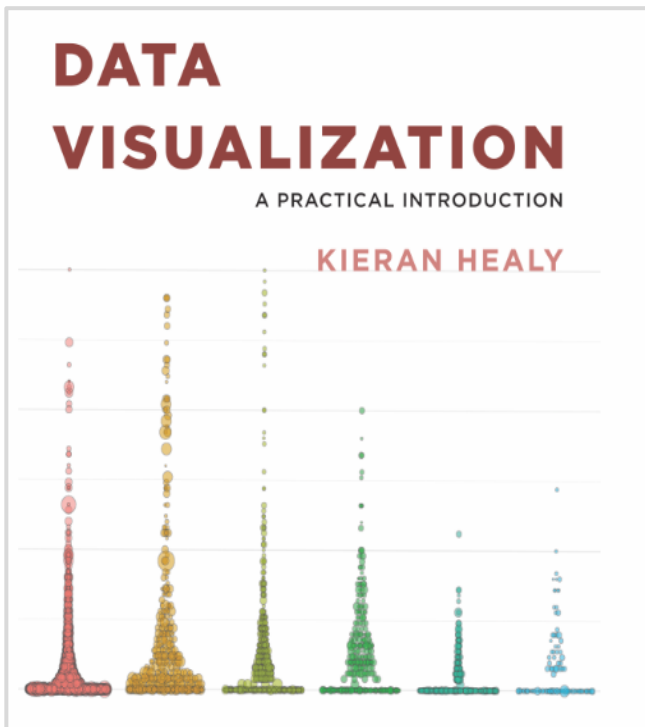
Trees, maps, and theorems

Effective communication for rational minds

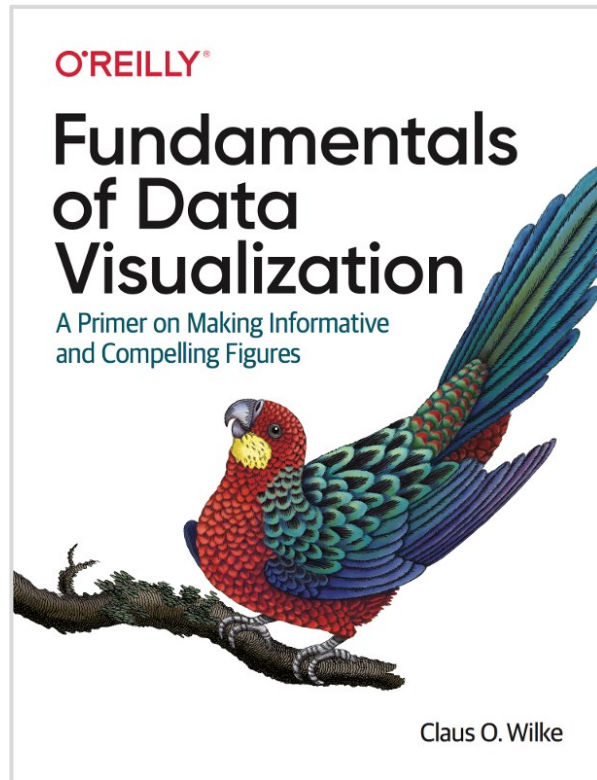
Jean-luc Doumont



<https://www.principiae.be/book/>



<https://socviz.co/>



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



# Wonderful Wednesdays 10

08-Dec-2020



EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY  
Representing Statistical Associations in Europe

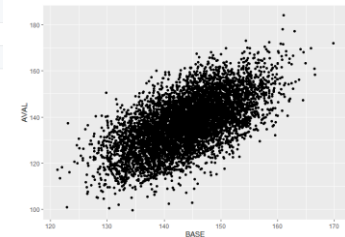
**An EFSPI/PSI VIS SIG initiative**

# Meta-analysis example data set

- The example simulated data set is based on **seven** phase III studies in Hypertension.
- A wide collection of baseline measurements are also included which can be explored to understand the patient populations within each trial, to search for potential subgroups or differential treatment effects, or even to develop prognostic or predictive risk models.
- For a detailed overview of the data set, please refer to the data dictionary provided:

<https://github.com/VIS-SIG/Wonderful-Wednesdays/tree/master/data/2020/2020-11-11>

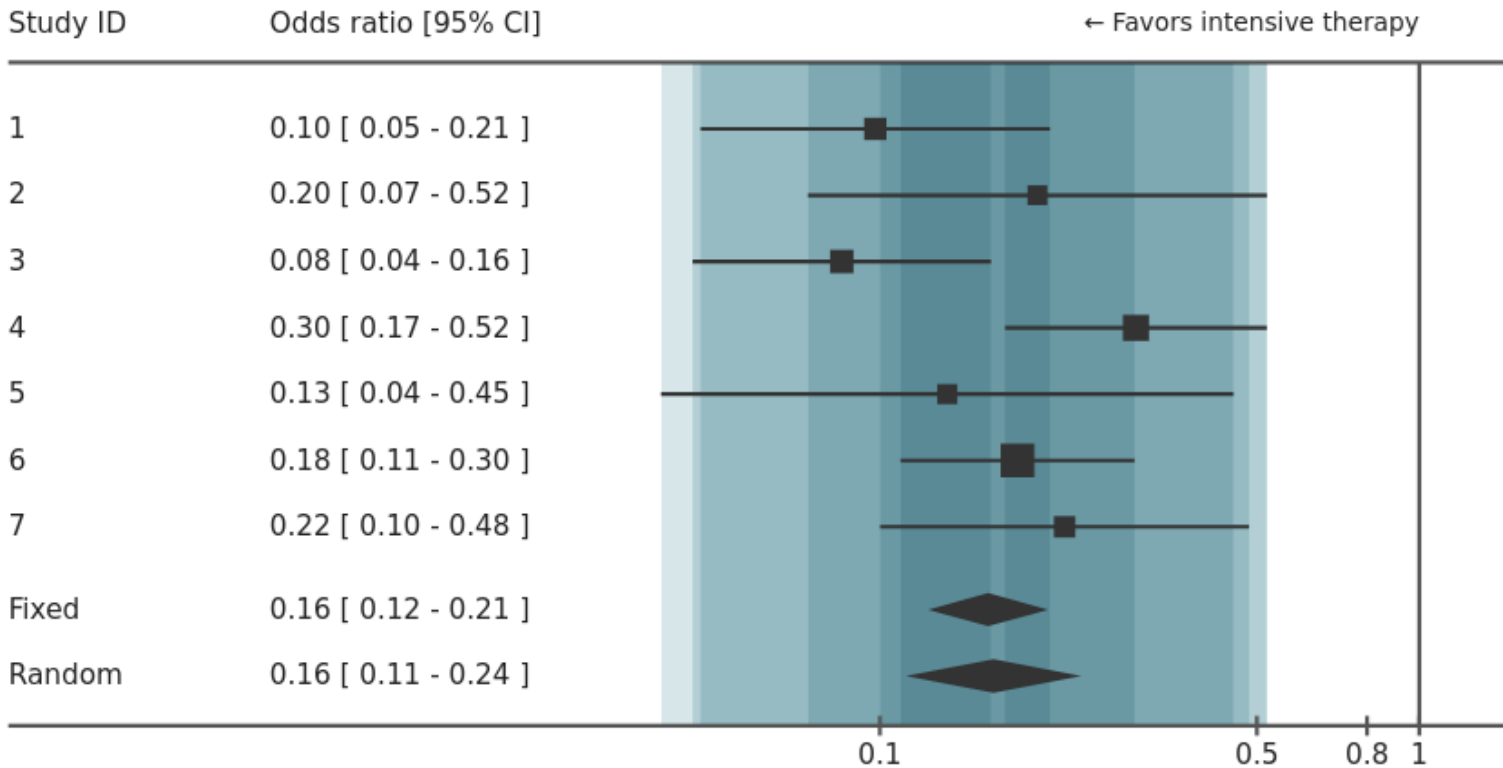
AGE	Age (years)	GGTSI	Gamma Glutamyl Transferase (U/L)
AGECAT1	Age Group 3: 75 years and older = TRUE	GLUCPSI	Glucose, Plasma, Fasting (mmol/L)
AGECAT1C	Age Group 3: 75 years and older = TRUE	GREGGR10	Regional stratification group
AGECAT1N	Age Group 3: 75 years and older = TRUE	HCT	Hematocrit
ALBSI	Albumin (g/L)	HDLSI	HDL Cholesterol (mmol/L)
BASE	Mean systolic blood pressure (mm Hg) measured at baseline	HDT	Phosphate (mmol/L)
BASOSI	Basophils (Absolute) (10E9/L)	HEIGHT	Height (cm)
BICARSI	Bicarbonate (mmol/L)	HGBSI	Hemoglobin (g/L)
BILSI	Bilirubin (umol/L)	KSI	Potassium (mmol/L)
BMI	BMI	LDLSI	LDL Cholesterol (Assayed) (mmol/L)
BUNSI	Blood Urea Nitrogen (mmol/L)	LPASI	Lipoprotein-A Protein (g/L)
CASI	Calcium (mmol/L)	LYMLES1	Lymphocytes/Leukocytes (%)
CHD10R1	10-year Coronary heart disease (CHD) risk category (High (>20%) , Me	LYMSI	Lymphocytes (Absolute) (10E9/L)
CHD10R1N	10-year CHD risk category (Numeric, 1 = Low, 2 = Medium, 3 = High)	MONLSI	Monocytes/Leukocytes (%)
CHOL_HDL	Ratio of Total Cholesterol / HDL	RACE	Race
CHOLS1	Cholesterol (mmol/L)	SBPCAT1C	Mean systolic blood pressure (mm Hg) at baseline (Category)
COUNTRY	Country indicator	SBPCAT1N	Mean systolic blood pressure (mm Hg) at baseline (Numeric)
CREATSI	Creatinine (umol/L)	SEX	Sex
EOSLES1	Eosinophils/Leukocytes (%)	TRIGFSI	Triglycerides (Fasting) (mmol/L)
EOSSI	Eosinophils (Absolute) (10E9/L)	URATESI	Uric Acid (umol/L)
ETHNIC	Ethnicity	WBCSI	Leukocytes (10E9/L)
GGTSI	Gamma Glutamyl Transferase (U/L)	WEIGHT	Weight (kg)
GLUCPSI	Glucose, Plasma, Fasting (mmol/L)		



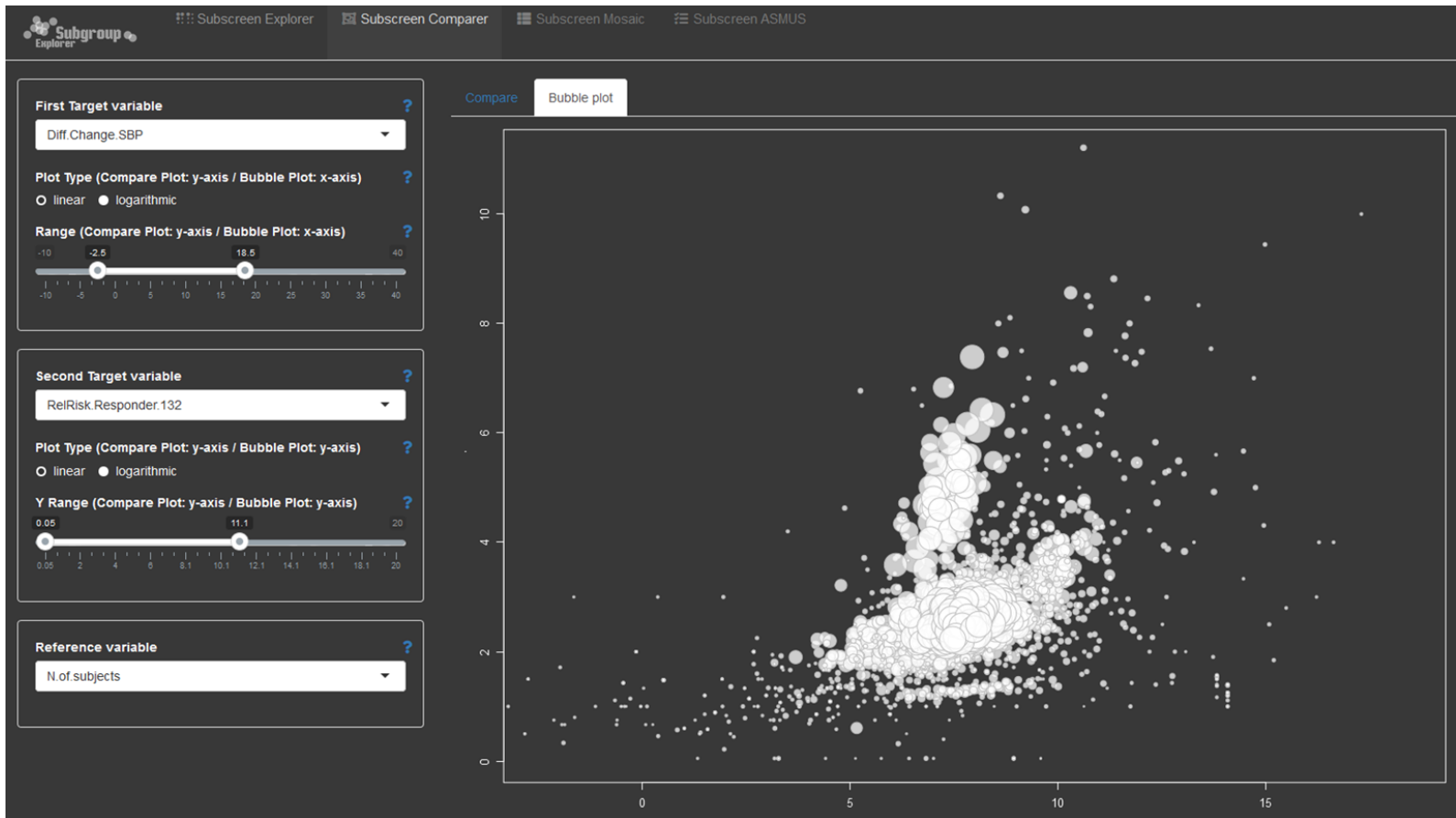


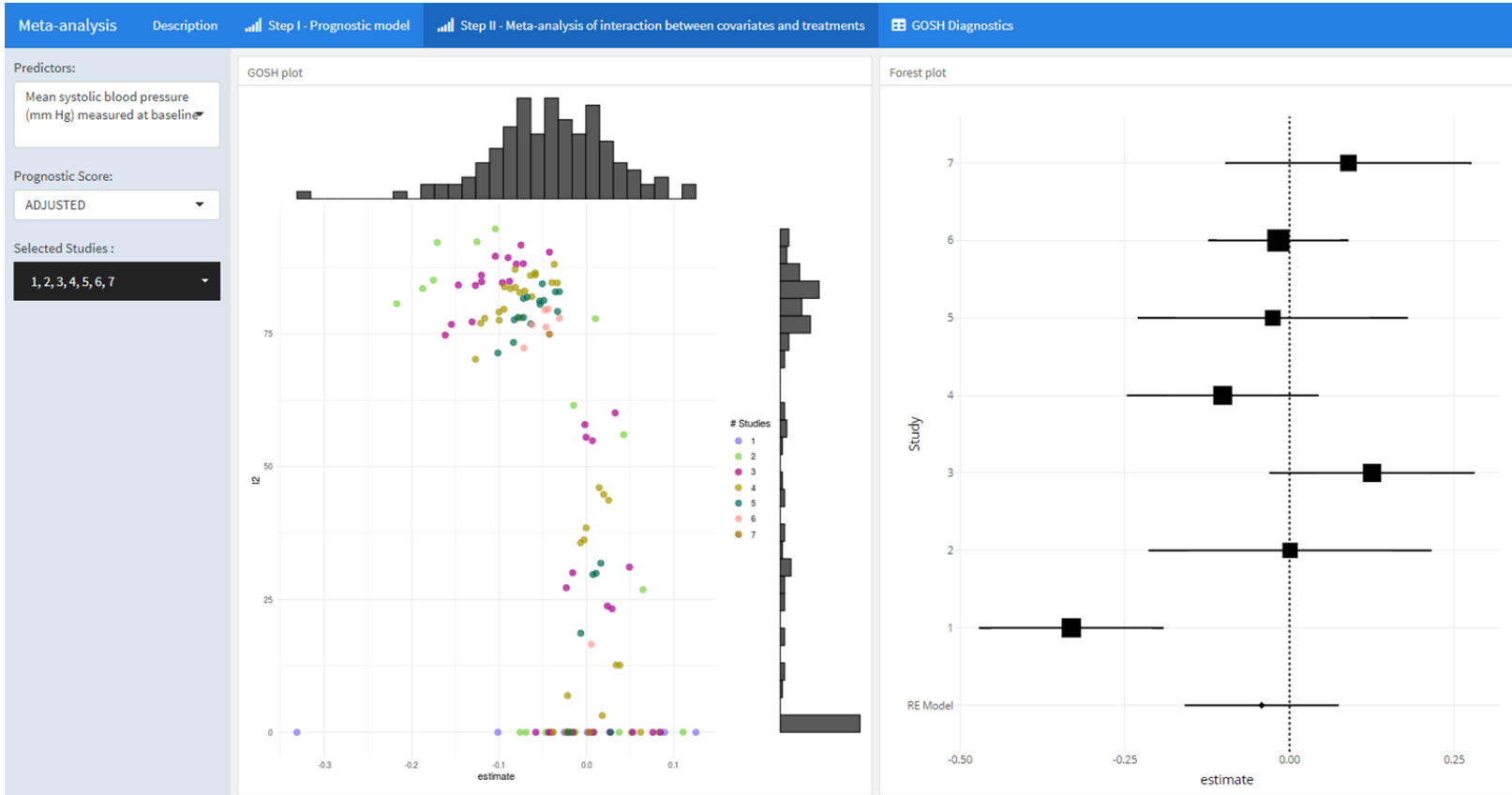
# Intensive antihypertensive therapy versus standard of care

Responder analysis - patients with controlled systolic blood pressure at 1 year ( $\leq 120$  mmHg)









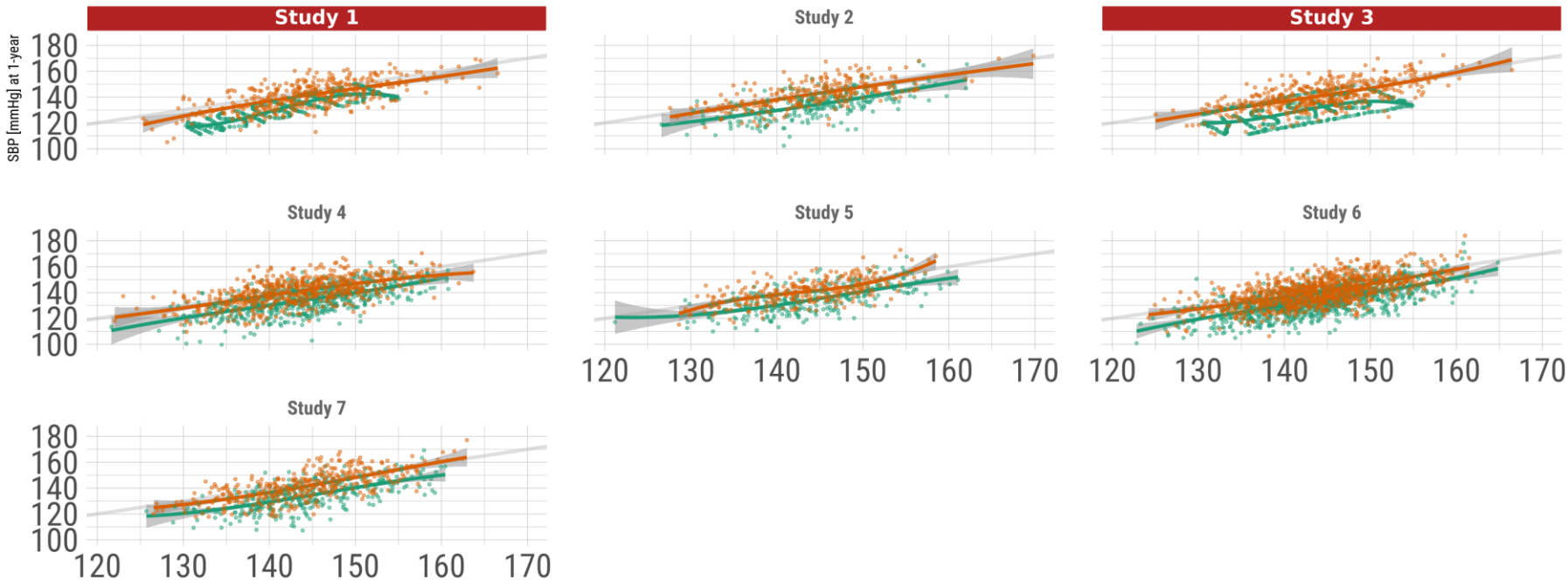
<https://figplot.shinyapps.io/WW20201209/>

# Meta-analysis example data set

- How data visualisation can be deployed to understand integrated data?
- **Key issues where data visualisation can help are around the investigation of whether studies can be combined due to study heterogeneity**
- This throws up questions such as:
  - What graphical tools can be used to assess heterogeneity?
  - What variables are prognostic or predictive of outcome?
  - Where can graphical methods provide general recommendations?

# Comparison of pre-post mean systolic blood pressure (SBP) measured at baseline and 1-year

Study 1 and 3 may have data quality issues - further investigation required



SBP [mmHg] at randomisation

Treatment — Intensive treatment — Standard of care

The by-treatment relationship also displayed using a cubic splines.  
y = x reference line also displayed.  
Data: BIG\_DATA\_PSI\_WW\_DEC2020.csv

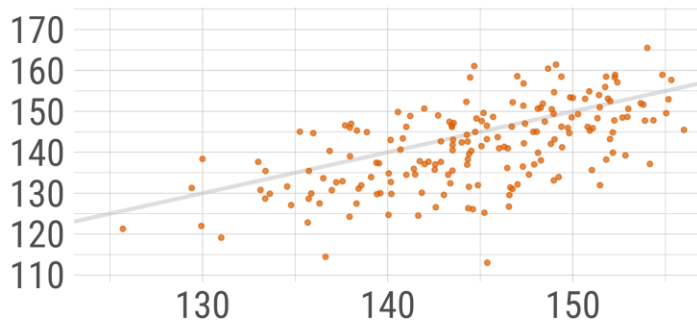
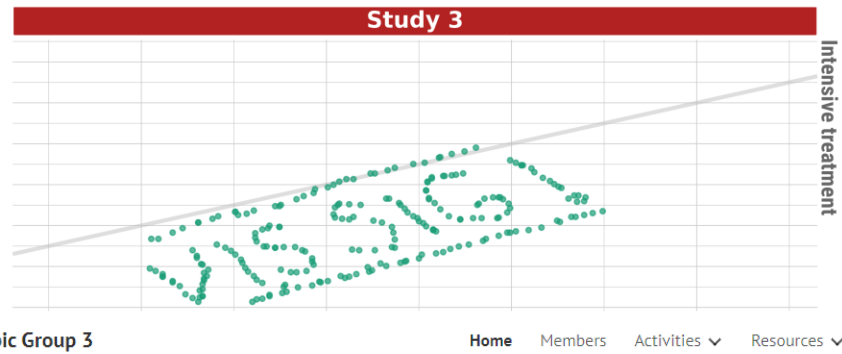
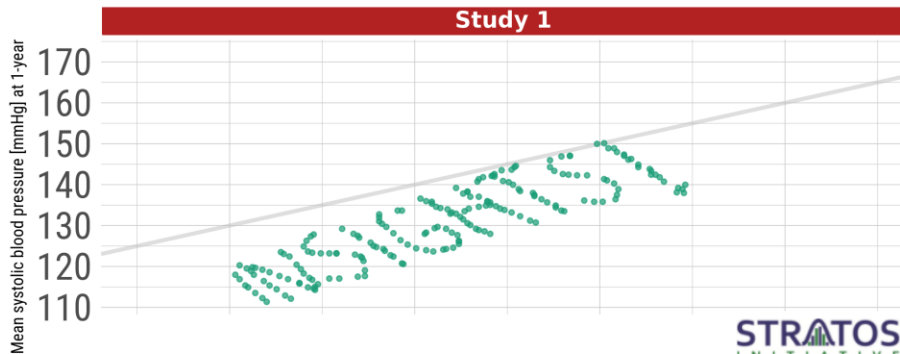
# The intensive treatment arm for study 1 and 3 displayed patterns of interest

It is always important to plot data many ways



## The intensive treatment arm for study 1 and 3 displayed patterns of interest

It is always important to plot data many ways



## Initial Data Analysis

### Aim

To improve awareness of Initial Data Analysis (IDA) as an important part of the research process and to provide guidance on conducting IDA in a systematic and reproducible manner.

### Key principle:

IDA should not touch the research question. Our group promotes initial data analysis (IDA) as a highly structured step in the data analysis process. For this purpose, we developed a framework for IDA and are creating tools to facilitate the IDA process.

# Effective data visualisation is effective communication

## Effective visualisations

- enable clear and impactful communication,
- elevate influence with stakeholders,
- facilitate informed decision making.

To help design effective visualisations, remember three principles:

- Purpose,
- Clarity
- Message

More here

- <http://www.stratos-initiative.org>
- <https://graphicsprinciples.github.io/>





<http://www.stratos-initiative.org>

**Thank you**