

Have a clear
purpose

Show the
data
clearly



Make the
message
obvious

STRATOS visualisation panel: virtual meeting update

Mark Baillie

November 9th, 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.”

Chair: Mark Baillie

Members: Tim Morris, Rachel Phillips, Willi Sauerbrei, Svenja Seide, Marc Vandemeulebroecke

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

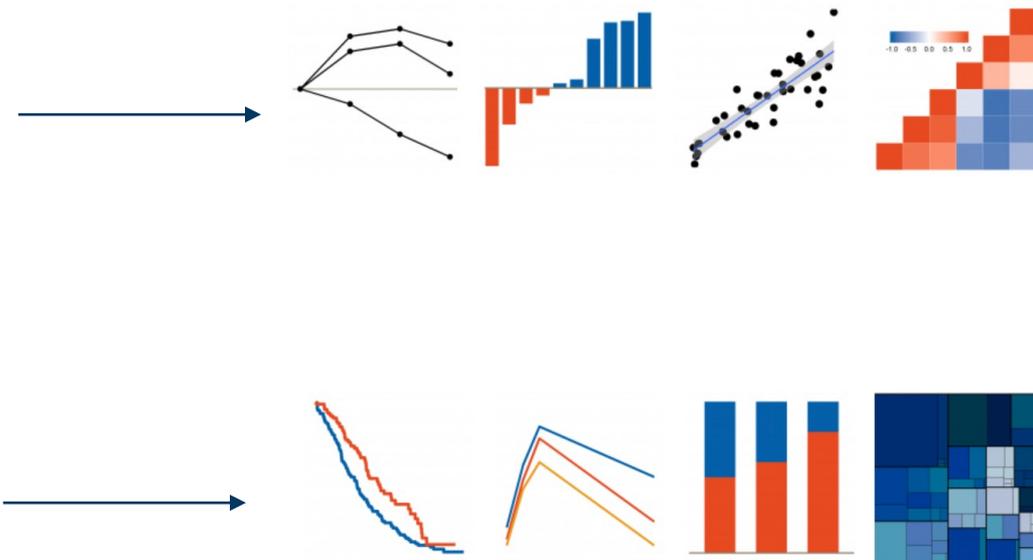
The logo for the STRATOS Initiative. The word "STRATOS" is written in a large, bold, dark blue sans-serif font. The letter "A" is stylized, with a green bar chart consisting of three vertical bars of increasing height integrated into its shape. Below "STRATOS", the word "INITIATIVE" is written in a smaller, green, all-caps sans-serif font, with wide letter spacing.

STRATOS
INITIATIVE

Effective visualisation is important throughout the workflow

Topic groups

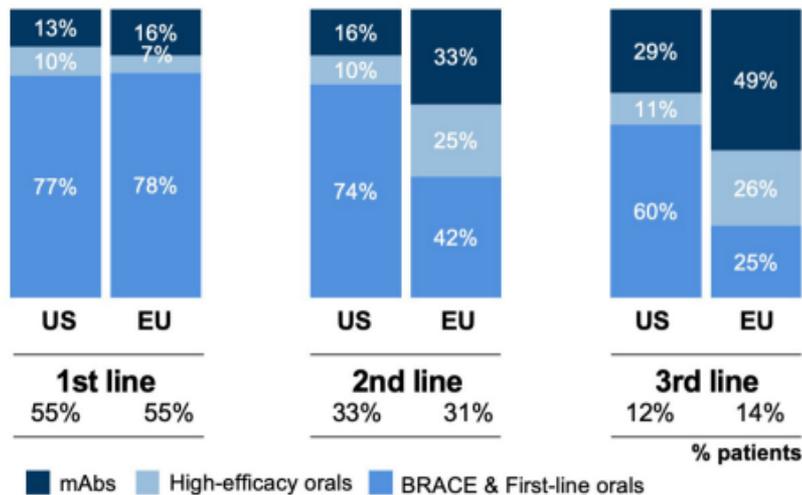
1	Missing data
2	Selection of variables and functional forms in multivariable analysis
3	Initial data analysis
4	Measurement error and misclassification
5	Study design
6	Evaluating diagnostic tests and prediction models
7	Causal inference
8	Survival analysis
9	High-dimensional data



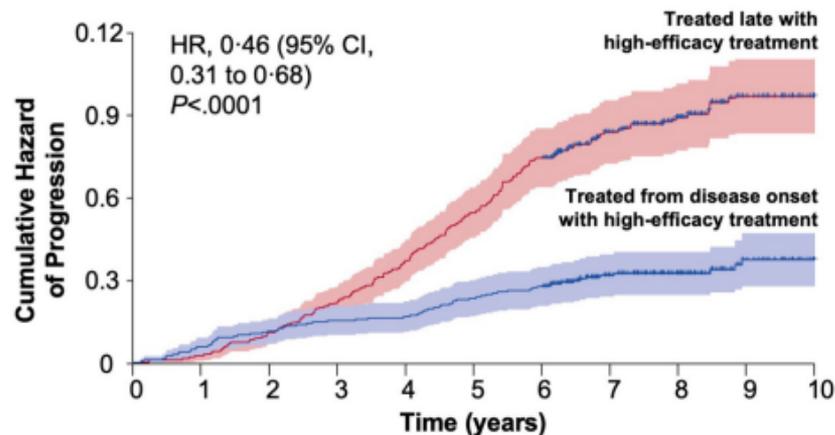
BRACE and first-line orals commonly used in early stages

Despite data showing high-efficacy treatments started early result in better outcomes

Use of disease-modifying treatments in MS¹



Cumulative hazard of CDP in patients with RRMS treated from disease onset versus late with high-efficacy treatment²



MAbs: Ocrevus®, Lemtrada®, Tysabri®; **High-efficacy orals:** Gilenya®, Mayzent®, Mavenclad®; **BRACE & First-line orals:** Interferons, Copaxone®, GA Gx; Tecfidera®, Aubagio® *High efficacy DMTs may include orals and MABs

1. Symphony APLD (Sep 2018-Aug 2019). , EU5 IPSOS Monitor 2019. 2. He A et al. Lancet Neurol. 2020;19(4):307–316. Retrospective analysis, measured from disease onset. Bold lines are cumulative hazard estimates and shaded areas are 95% CIs. CDP, confirmed disability progression; CI, confidence interval; HR, hazard ratio; RMS, relapsing multiple sclerosis.

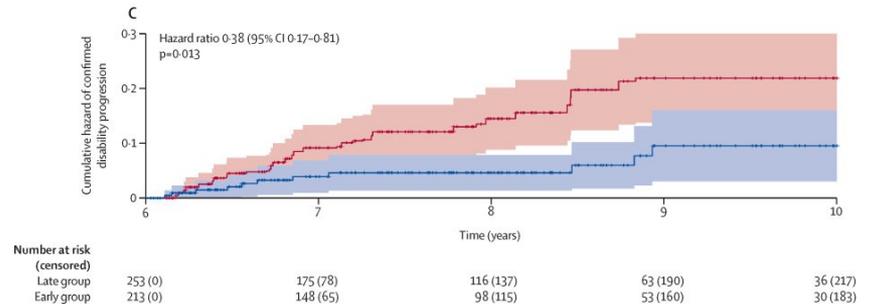
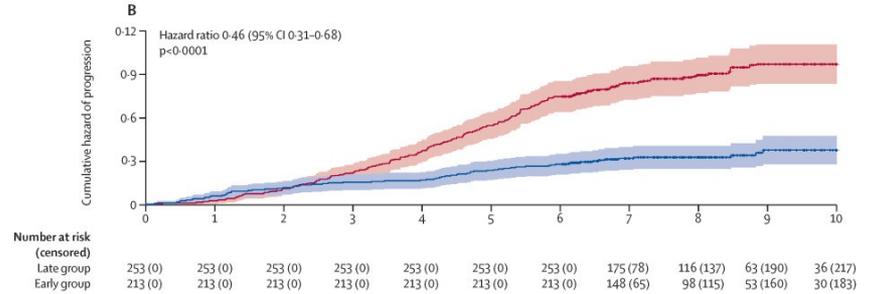
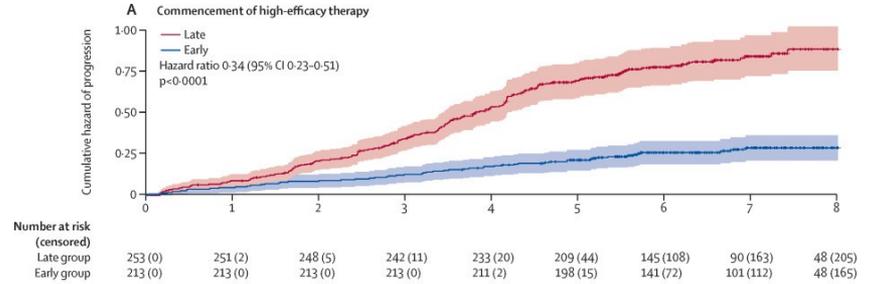


Articles

Timing of high-efficacy therapy for multiple sclerosis: a retrospective observational cohort study

Anna He MBBS ^a, Bernd Merkel PhD ^a, James William L Brown PhD ^{a, b, c}, Lana Zhovits Ryerson MD ^d, Ilya Kister MD ^d, Charles B Malpas PhD ^a, Sifat Sharmin PhD ^a, Dana Horakova MD ^{e, f}, Eva Kubala Havrdova PhD ^{e, f}, Tim Spelman PhD ^g, Guillermo Izquierdo MD ⁱ, Sara Eichau MD ⁱ, Maria Trojano MD ^j, Alessandra Lugaresi MD ^{k, l}, Raymond Hupperts MD ^m, Patrizia Sola MD ⁿ, Diana Ferraro MD ⁿ, Jan Lycke PhD ^o, Francois Grand'Maison MD ^p, Alexandre Prat MD ^{q, r}, Marc Girard MD ^{q, r}, Pierre Duquette MD ^{q, r}, Catherine Larochele MD ^{q, r}, Anders Svenningsson MD ^h, Thor Petersen MD ^s, Pierre Grammond MD ^t, Franco Granella MD ^u, Vincent Van Pesch MD ^v, Roberto Bergamaschi MD ^x, Christopher McGuigan MD ^{y, z}, Alasdair Coles PhD ^b, Jan Hillert MD ^g, Fredrik Piehl MD ^g, Helmut Butzkueven PhD ^{aa, ab}, Tomas Kalincik PhD ^{a, ac, ad} on behalf of the MSBase study group [†]

<https://www.sciencedirect.com/science/article/pii/S1474442220300673?via%3Dihub#fig3>



Visualisation Panel Updates

- Biometric Bulletin article 2020
- Support publication panel - provide review and support of STRATOS publications
- Presentation ROeS Sept 2021
- Collaboration with TG2-TG3
 - Regression without regrets (cross-sectional)
 - Leads: G. Heinze, M. Baillie, M. Huebner, a TG2-TG3 collaboration
- Workshops - M. Baillie and M. Vandemeulebroecke.
 - Developed hands on workshop and rolled out to 1,000 statisticians, clinicians and other quantitative scientists at Novartis.
 - Active learning – pen, paper and gray matter
 - Designed to be face to face, 3 hours, but now redeveloped to run virtually
 - Early 2021 invited presentation to ASCPT Pharmacometrics & Pharmacokinetics
 - 2022 Will present a version of the workshop at IBC
 - <https://members.biometricsociety.org/ibc2022/events/showcase>
- Related initiatives - R. Phillips & M. Baillie
 - Founding members of active PSI special interest group on visualization
 - <https://www.psiweb.org/sigs-special-interest-groups/visualisation/welcome-to-wonderful-wednesdays>

Principles for EVC

Graphics Principles Cheat Sheet v1.1

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.

Planning

Clearly identify the purpose of the graph, e.g. to deliver a message or for exploration?

Identify the quantitative evidence to support the purpose

Identify the intended audience (specialists, non-specialists, both) and focus the design to support their needs

Identify the intended audience (specialists, non-specialists, both) and focus the design to support their needs

Identify the intended audience (specialists, non-specialists, both) and focus the design to support their needs

Adapt the design to space or formatting constraints (e.g. critical report, slide deck or publication)

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 & placement to 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. Use contrasting colors to draw attention

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 scheme, and spacing. Use images with similar styles

Effectiveness Ranking

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

Volume Color Hue Depth: 3d position Color intensity Area Shape or Angle Length Position on Unaligned scale Position on Common scale



Least accurate → Most accurate

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

Selecting the right base graph

Consider if a standard graph can be used by identifying suitable designs based on the: (i) **purpose** (i.e. message to be conveyed or question to answer) and (ii) **data** (i.e. variables to display).

Deviation	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Magnitude
Chg. from baseline	Scatter plot	Horizontal bar chart	Boxplot	Kaplan Meier	Stacked bar chart	Vertical bar chart

Facilitating Comparisons

Proximity improves association

Place labels next to data instead of using legends

Group together elements that are compared directly

Do not use color to differentiate 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.

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

Use reference lines and other visual anchors.

Use reference lines and other visual anchors.

Reduce mental arithmetic

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

Use reference lines and other visual anchors.

Implementation Considerations

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

Aspect ratio can influence interpretation. Aim for a 45 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 scale or % change).

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

Plot data and inferences to support stories about nodes.

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 such in order to avoid confusion.

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 2x larger than all other labels to make them more prominent).

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

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 set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

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 set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

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 set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

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 set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

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 set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

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

Good graph checklist

Clear Communication

Is the message of the graph as clear as possible?

Is it easy for someone unfamiliar with the data to interpret the graph?

Are the patterns/relationships easily identified?

Is the graph tailored to its primary purpose and audience?

Is the correct graph type used?

Facilitating Comparisons

Are elements to be compared grouped together?

Are labels placed next to data instead of in legends?

Have categories been ordered for easy comparison?

Can the plot be read without doing mental calculations?

Are the estimates of interest plotted (e.g. mean differences with confidence intervals)?

Color for emphasis or distinction

Are graphical elements displayed in a dark color on a light background?

Are grid lines drawn with a thin line and a light color such as grey?

Are colors used sparingly (e.g. max 3)?

Do all elements in the graph have a purpose (e.g. colors, textures, grid lines)?

Are the same colors used to mean the same thing in a series of graphs?

Implementation Considerations

Are multiple panels plotted on the same scale?

Are logarithmically distributed variables plotted on a log scale?

Are common baselines used wherever possible?

Does the orientation of the axes aid interpretation?

Does the aspect ratio allow the reader to see variations in the data?

Are data across a disconnected time scale kept disconnected?

Are data spaced proportionally to the actual time interval (instead of according to visual number)?

Are data and inferences plotted to support stories about models?

Are number of patients by group reported if this adds context?

Legibility and Clarity

Can all graphical elements be seen?

Does the graph have a clear title, axis labels, annotations and data units?

Can the font be read without eye strain or effort?

Are sans-serif fonts used?

Do text sizes have correct hierarchy (big to small, main text to subtext)?

Are the elements of the graph clearly labeled (e.g. points, error bars, lines, shaded regions)?

Are labels oriented horizontally where possible?

© 2017 Novartis



Resources

Books:
 E.R. Tufte. The visual display of quantitative information. Cheshire, Conn.: Graphics Press, 2001.
 Cleveland, W.J. and McGill, Robert. Graphical perception: theory, experimentation and application to the development of graphical methods. JASA, Vol. 79, No. 387, pp. 531 – 554, 1984.
 S. Few. Show Me The Numbers: Designing Tables and Graphs to Enlighten (2nd Edition). Burlington, CA: Analytics Press, 2012.
 D. M. Wong. The Wall Street Journal Guide to Information Graphics: The Dots and Dots of Presenting Data, Facts, and Figures. December 16, 2013.
 J. Dourson. Titles, Maps, and Themes: Effective communication for rational minds. PRINCIPAL, N. B. Robbins. Creating More Effective Graphs. Chart House.

Online resources:
 https://www.pinterest.com/ (St. Paul)
 https://www.researchgate.net/publication/ (E. Tufte)
 https://www.researchgate.net/ (St. Paul)
 https://www.researchgate.net/publication/ (A. Gelman)
 https://www.researchgate.net/publication/ (A. Gelman)
 https://www.researchgate.net/publication/ (A. Gelman)
 https://www.researchgate.net/publication/ (A. Gelman)

Authors
 Alison Mackenzie, Mark Bellis, Bastien Magnussen, Julie Jones, Marc VanDerMeulen



Resources, where to find out more?

Graphics principles

Tutorial

Effective Visual Communication is essential in every step of a quantitative workflow, from scoping to execution and communicating results and conclusions. The goal of the [tutorial](#) is to convey this competency.

In the [tutorial](#) we posit three laws of Effective Visual Communication for the quantitative scientist: have a clear purpose, show the data clearly, and make the message obvious. The [tutorial](#) provides more granularity on these laws; it covers the [Graphics Principles Cheat Sheet](#); and it discusses the application of both in four [Case Studies](#).

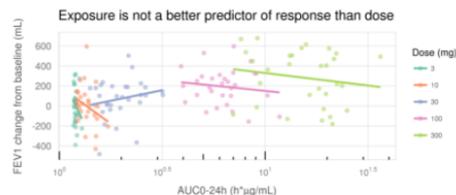


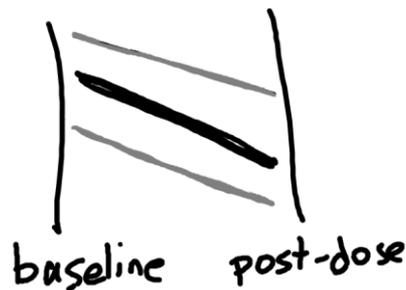
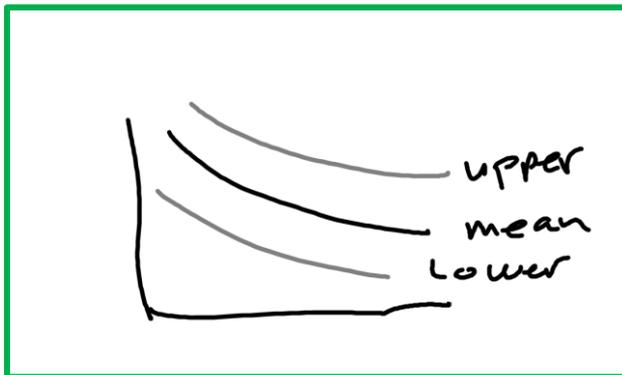
Figure 1: Example visualization from the tutorial

Additional resources:

- [Tutorial: https://graphicsprinciples.github.io/](https://graphicsprinciples.github.io/)
- [Exploratory graphics: http://go/xGx](http://go/xGx)
- [Graph gallery: http://go/graphgallery](http://go/graphgallery)
- Cheat sheet: [GraphicsCheatSheet](#)
- [Paper on effective visual communication](#)
- [Video on the three principles](#)
- Presentation checklist <http://go/presentationchecklist>
- [EVC sharepoint](#)
- [Wonderful Wednesdays](#)

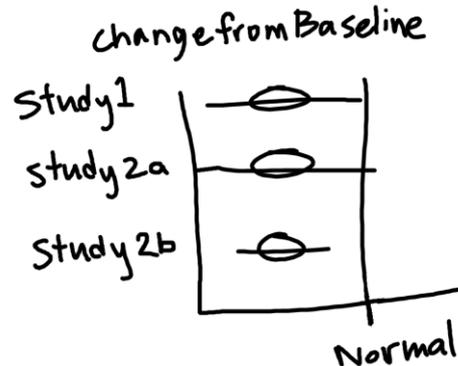
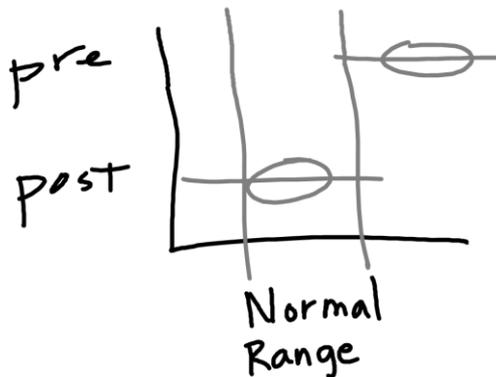
2nd principle - select the appropriate graph

- 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



Play the EVC Cartoon

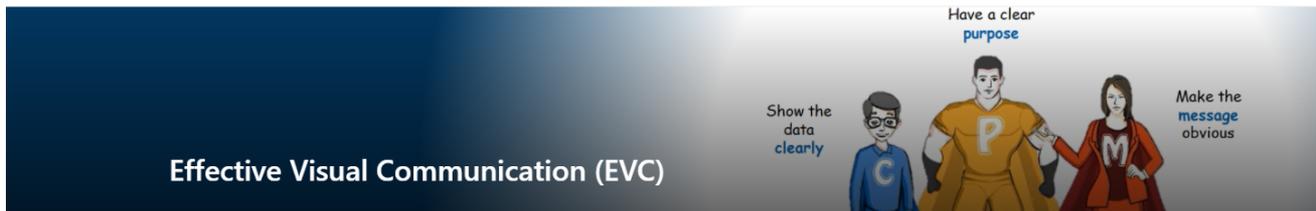
Have a clear **purpose**



▶ 🔊 1:49 / 3:54



Sharepoint: collection of resources



Scientists are generating more and more data every day. Our job as quantitative scientists is not only to analyze this data, but also to communicate our analyses in an efficient, effective, and impactful way. As Chambers et al. so elegantly stated:

“There is no single statistical tool that is as powerful as a well-chosen graph” - Chambers et al 1983

Effective visualizations may appear obvious, but creating them **requires skill**, patience, trial and error. Whether you are just beginning your career, or have many years of experience under your belt, **the following resources can help you** to develop effective data visualizations.



Internal Resources

- **Graph Gallery** (go/graphgallery) - dynamic resource for graphics and corresponding code, displayed in a web browser. Users can browse through samples or search for certain tags (e.g. SAS, R) or keywords (e.g. "Survival", "Safety"). Registered users can also rate and comment on individual graphs, participate in discussion forums, upload new graphs (along with code), and request help from graphical experts.
- **Graphics Principles Cheat Sheet** (go/graphcheatsheet) - reference sheet summarizing essential graphical principles on a single page, from selecting the right base graph, to facilitating comparisons (proximity, alignment), effective use of color (minimal, only for

Yammer: community of practice



Effective Visual Communication

NEW CONVERSATIONS

ALL CONVERSATIONS

FILES

SEARCH

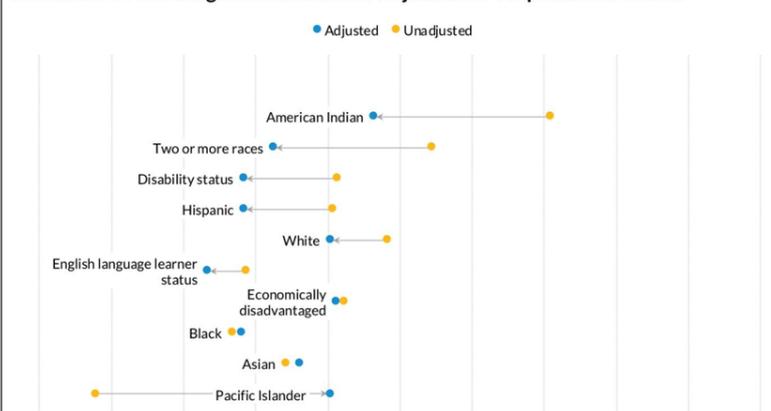


Magnusson, Baldur [Follow](#) – Friday at 12:32 PM

Making better data charts: From communication goals to graphics design

From Andrew Gelman's blog, [here](#) is an interesting post on the topic of making better graphs. Among other things, the article touches on the importance of knowing the purpose of the graph and having a clear message, and also gives various tips on clarity (including a handy link to the [directlabels](#) package in the comments).

Declines in Prekindergarten Enrollment Adjusted for Prepandemic Trends



PINNED

[Add](#)

- [Advanced Visual Analytics \(...\)](#)
- [Graphics Cheatsheet \(go/g...\)](#)
- [Graph Gallery \(go/graphga...\)](#)
- [EVC video \(go/graphvideo\)](#)

RELATED GROUPS

+ [Add a related group](#)

- [R User Community](#)
- [Shiny User Community](#)
- [CD&A Clinical Development & An...](#)

ACCESS OPTIONS

[Subscribe to this group by email](#)



Wonderful Wednesdays 10

08-Dec-2020



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

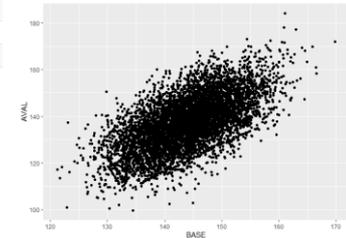
An EFSPI/PSI VIS SIG initiative

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:

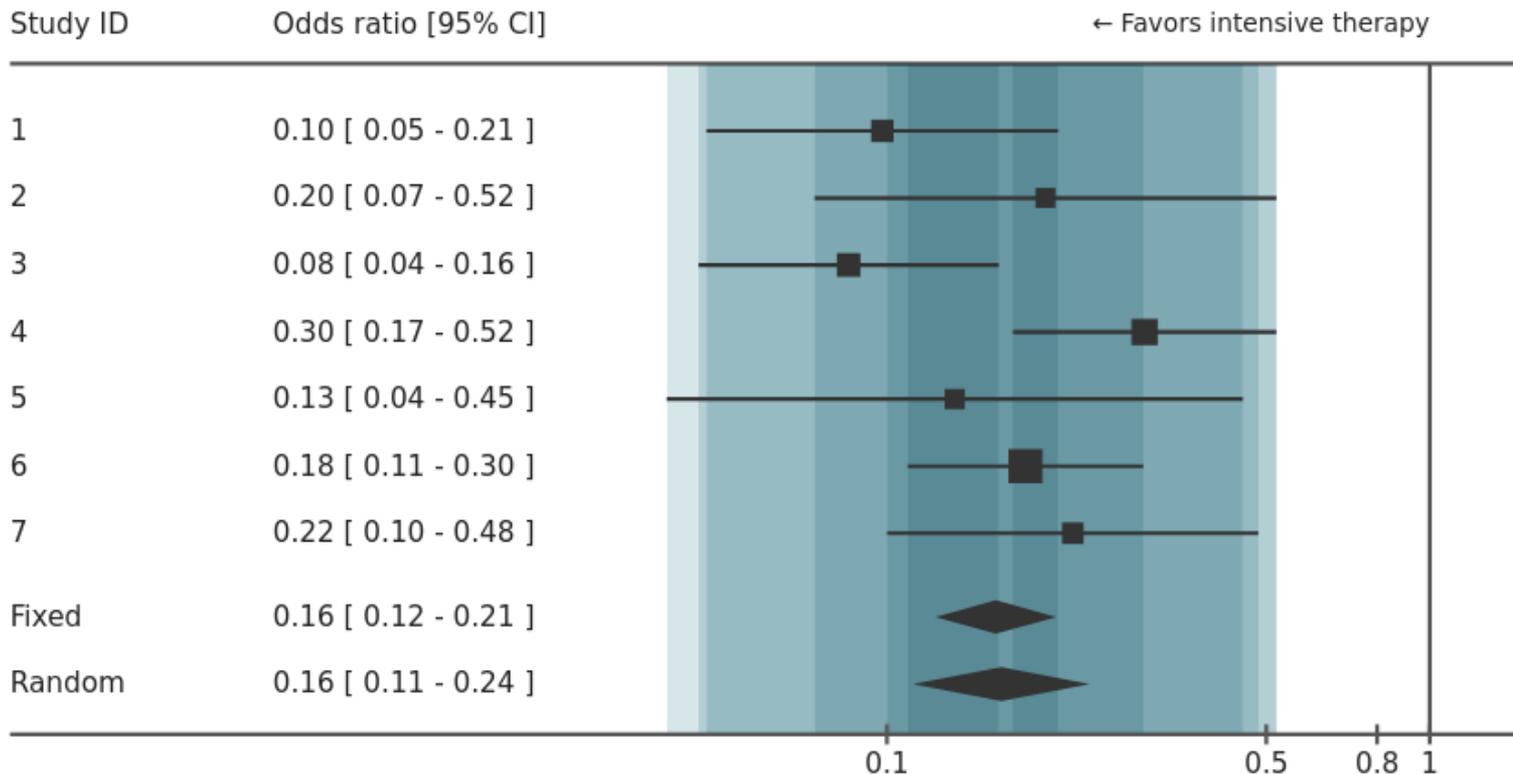
<https://github.com/VIS-SIG/Wonderful-Wednesdays/tree/master/data/2020/2020-11-11>

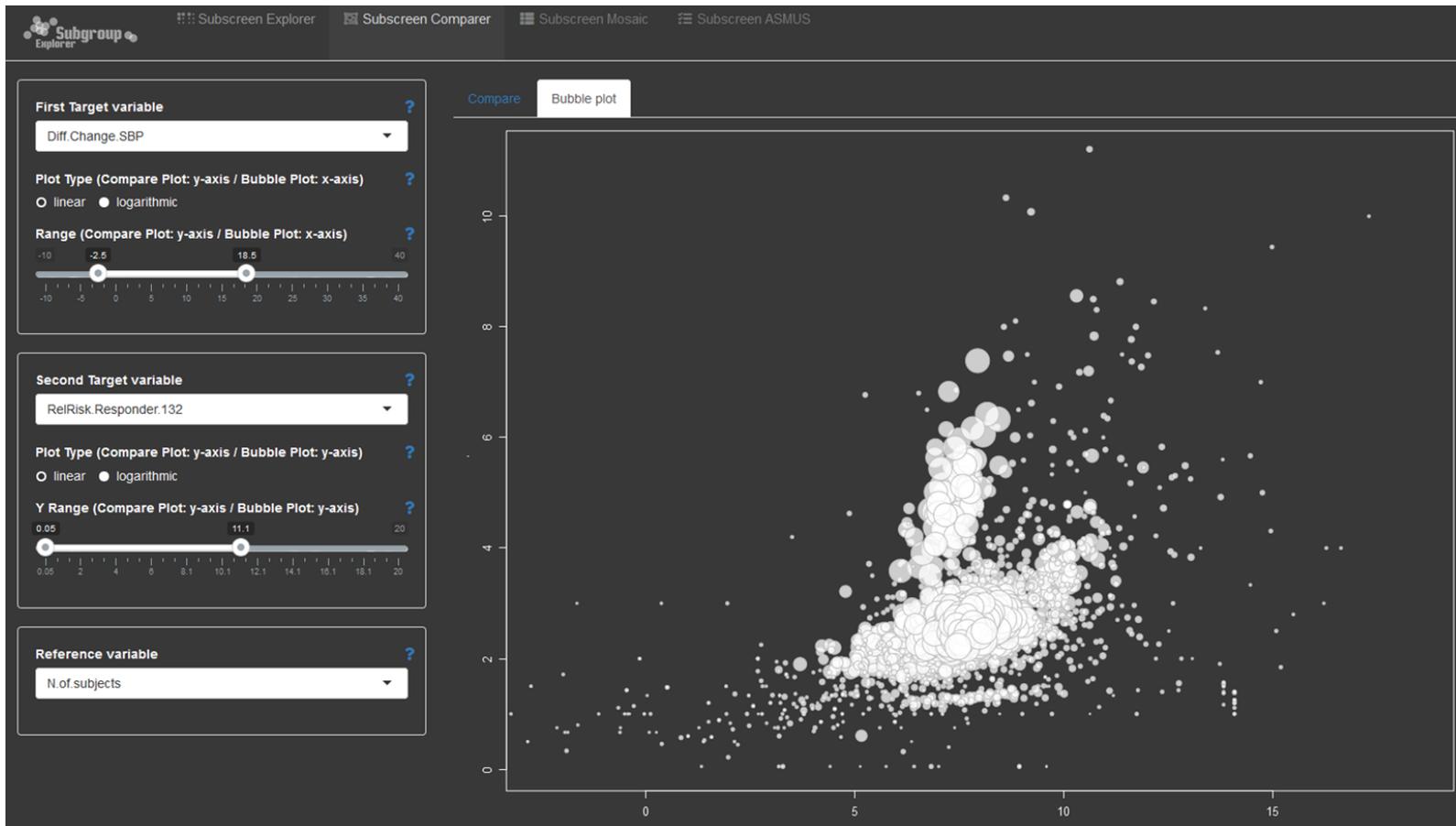
AGE	Age (years)	GGTSI	Gamma Glutamyl Transferase (U/L)
AGECAT1	Age Group 3: 75 years and older = TRUE	GLUCPSI	Glucose, Plasma, Fasting (mmol/L)
AGECAT1C	Age Group 3: 75 years and older = TRUE	GREGGRI0	Regional stratification group
AGECAT1N	Age Group 3: 75 years and older = TRUE	HCT	Hematocrit
ALBSI	Albumin (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)		

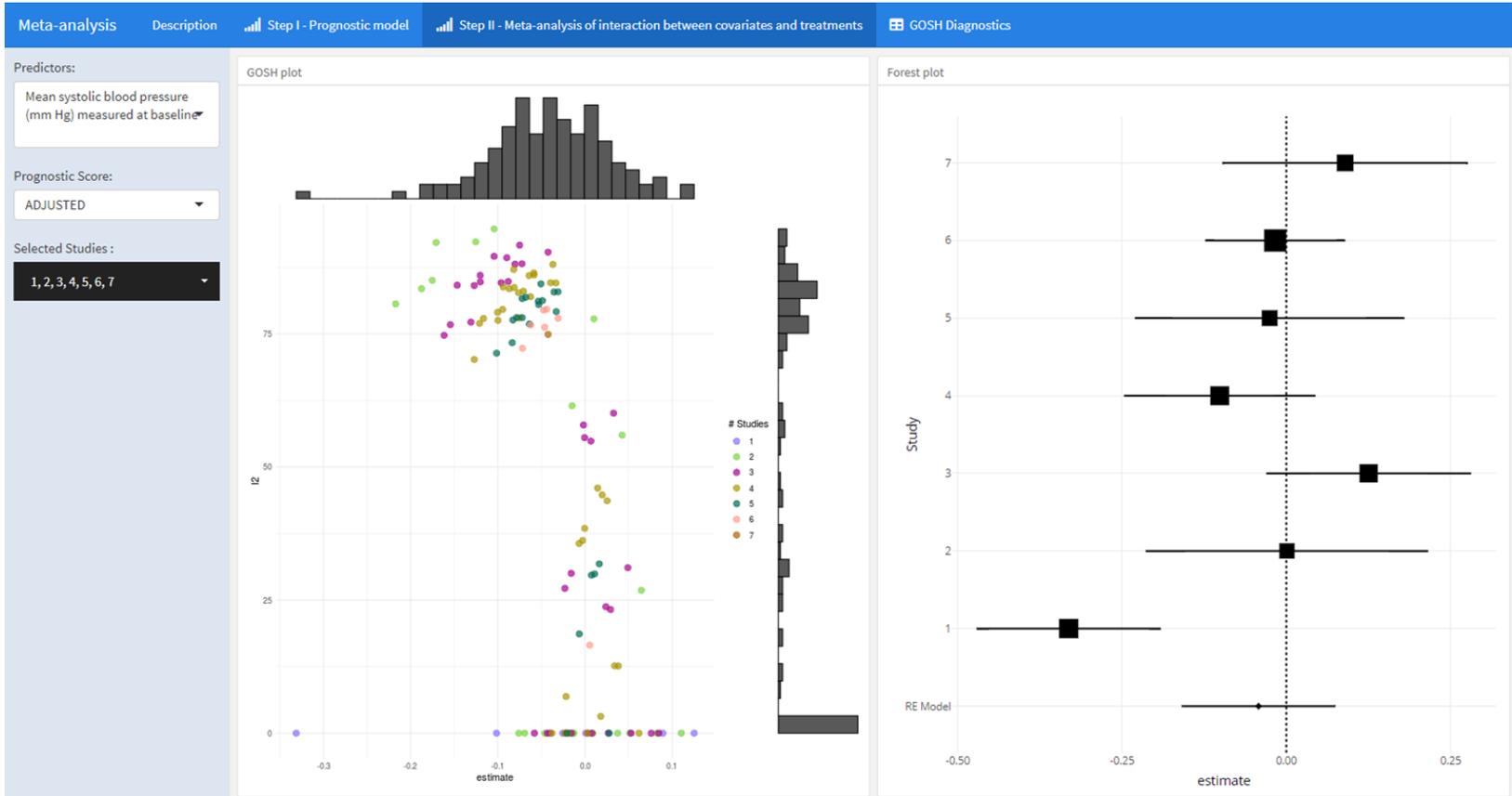


Intensive antihypertensive therapy versus standard of care

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







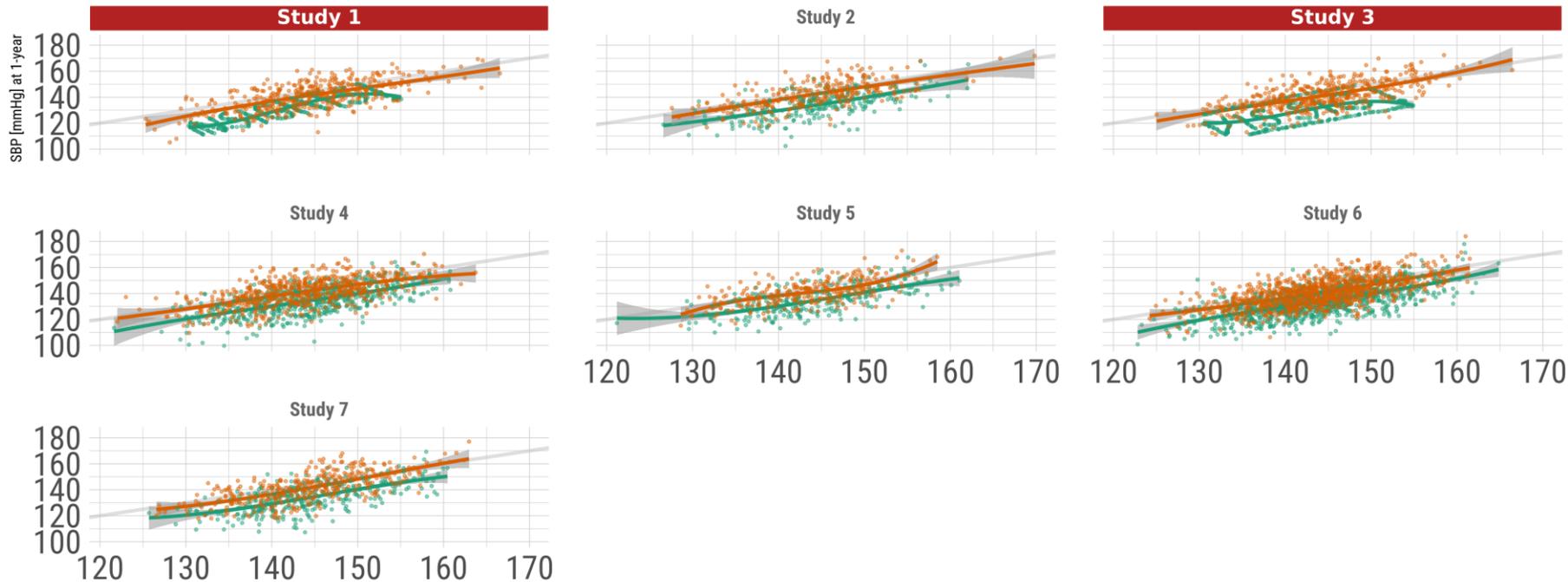
<https://figplot.shinyapps.io/WW20201209/>

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



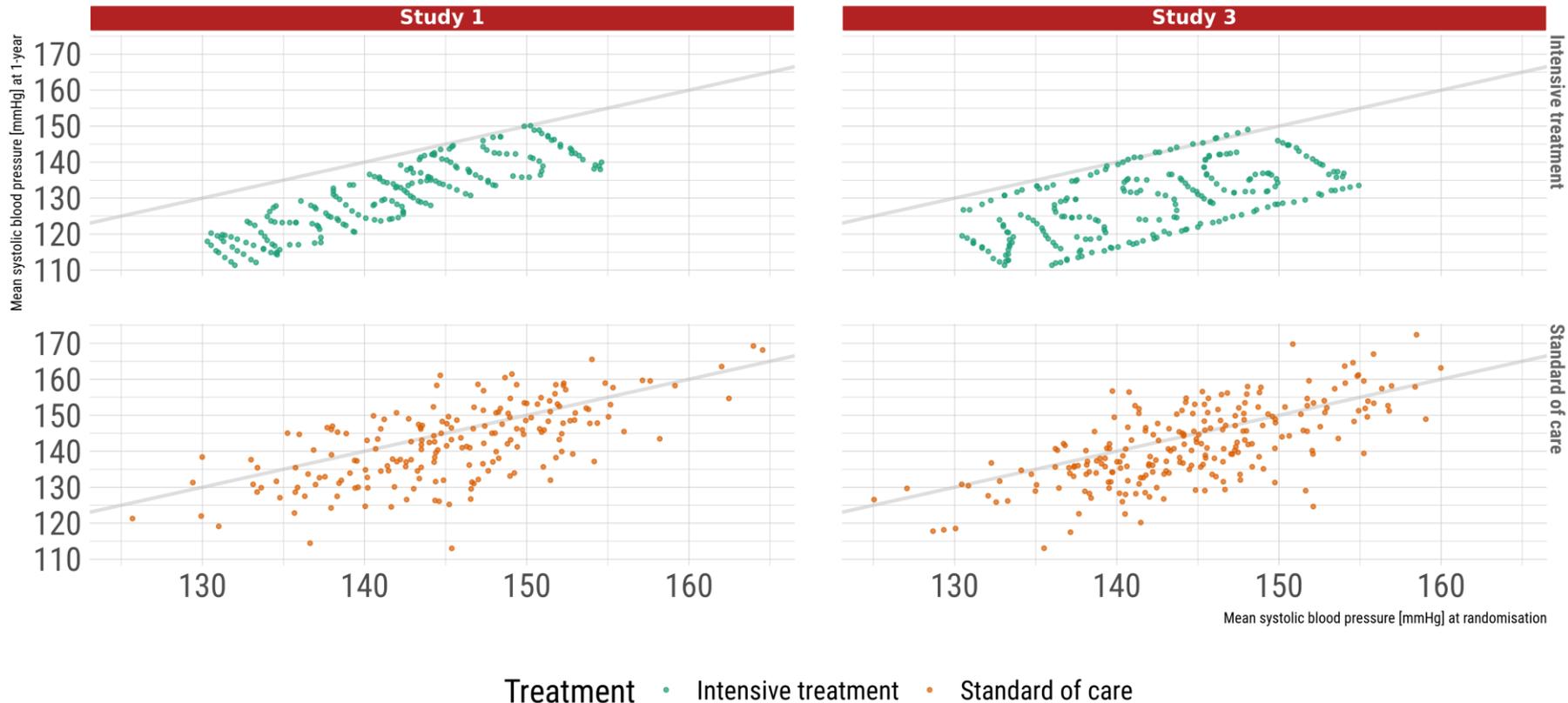
SBP [mmHg] at randomisation

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

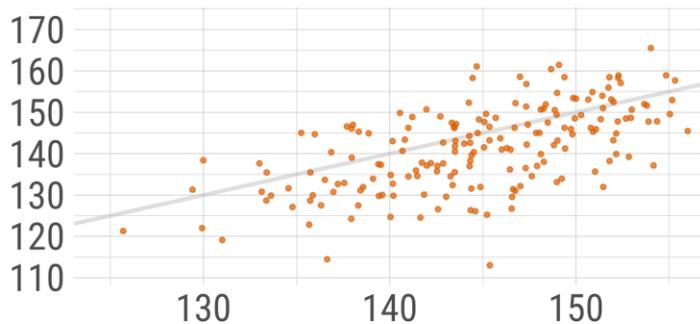
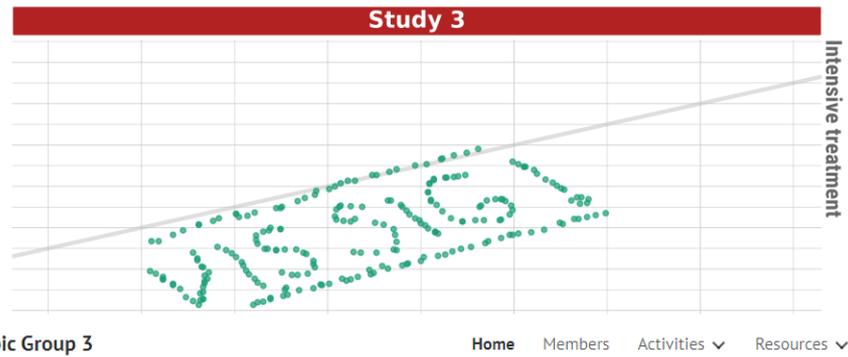
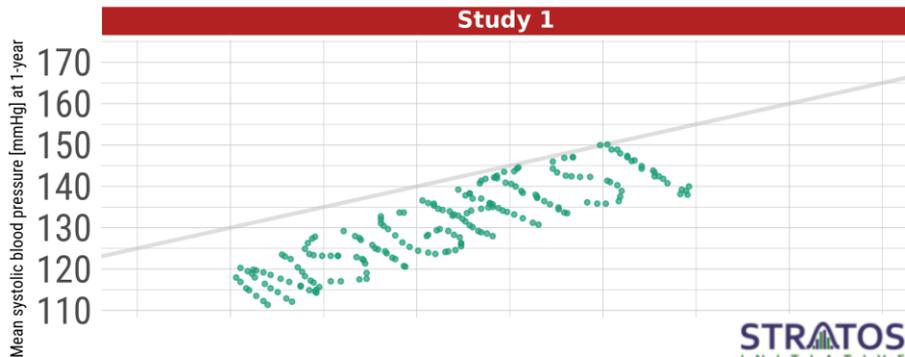
The intensive treatment arm for study 1 and 3 displayed patterns of interest

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.

Discussion

- Would like to identify co-chairs, help with activating the panel
- General thoughts or input from the wider group on future direction
- Questions?