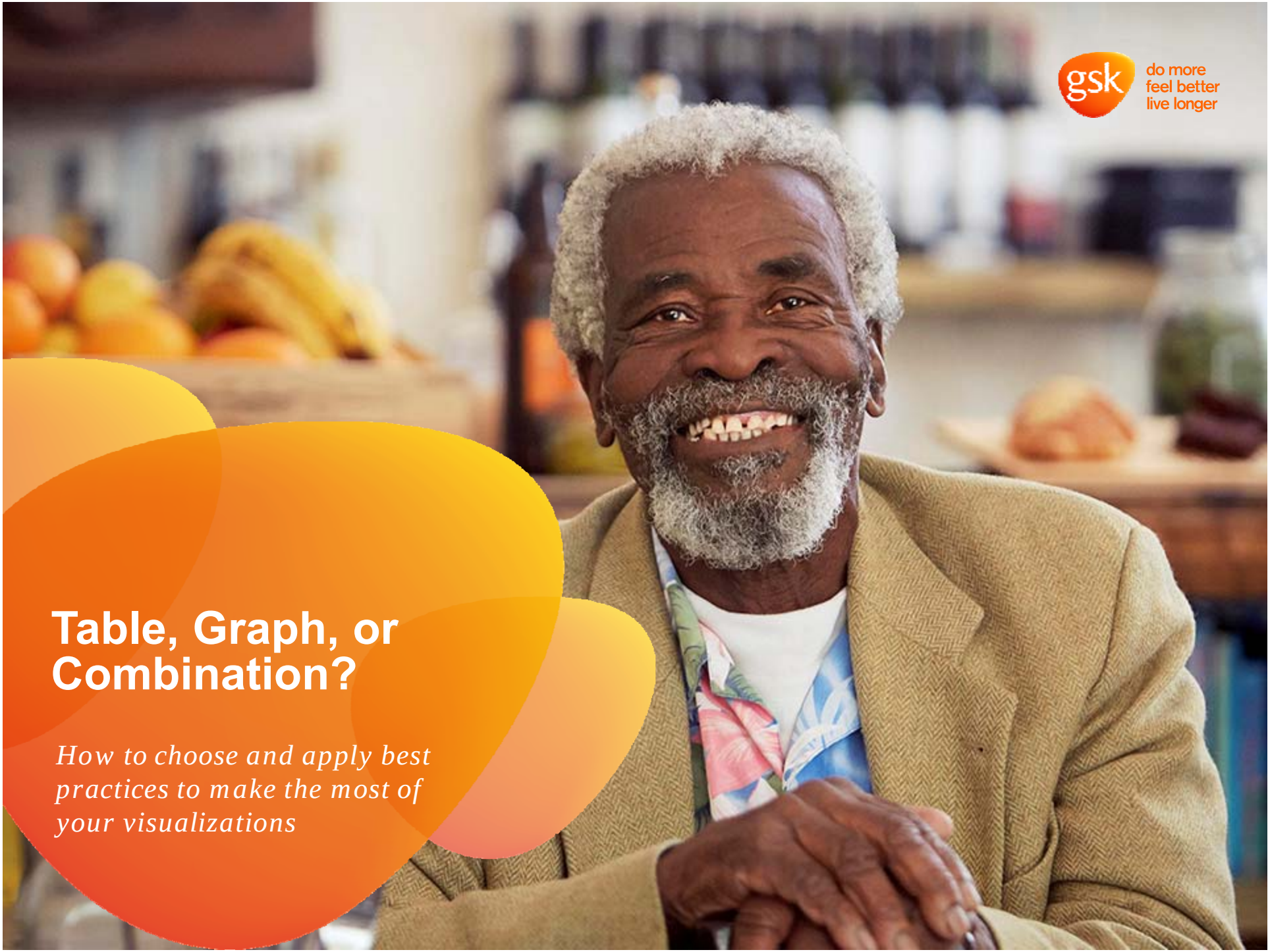


Table, Graph, or Combination?

How to choose and apply best practices to make the most of your visualizations

Table, Graph, or Combination?

How to choose and apply best practices to make the most of your visualizations

– **2016 PhUSE SDE**
– Frenchtown, NJ
– 2016-08-04

– **Author:**
– Daniel Schramek
– Programming Manager
– Oncology Therapy Area
– GlaxoSmithKline
– Collegeville, PA

Agenda

- Why data presentation is important
- When to choose text, tables or figures
- How to choose the right figure
- Good design principles

Why data presentation is important

Why data presentation is important

Purposes of data displays

- To efficiently and effectively translate key clinical messages from the data
- Communicates information for decision makers
- Describe, archive and list raw/collected data (eCRF, vendor data)
- Summarize, archive and list analysis data
- Recognize and reveal patterns, differences, significance, etc with data

Why data presentation is important

Who are decision makers?

- Upper Management, Regulators, Healthcare Providers, Patients
- Most likely not to have formal training in statistics
  - May be naïve to the details of the study design, objectives or history of the disease
  - May be less quantitatively and analytically skilled than the data analyst
  - May need more assistance in correctly interpreting the observed results

Why data presentation is important

Visual displays enhances cognition

- Data displays can organize information by indexing it spatially – items 'closer' together are similar or related to each other
  - Two nearby cities on a map are closer together
  - In a scatter plot, items plotted closer together are related
  - Other ways to represent 'closeness/relatedness' – color grouping line types, connected lines, etc.
  - Lining up columns of a table by related timepoints for two different treatments
- Offloading of cognition
  - Data displays are stored externally on some visualization media, freeing up brain power to focus on interpretation instead of manipulation
  - Data displays can enable complex computations to be replaced by simple pattern recognition processes
  - The ability to interact, filter or rotate displays frees us from having to do it mentally
- The design of the data display will directly affect the ease or difficulty of interpretation of the intended message

When to use text, tables or figures

When to use text

- Use text for simple or headline results which can be stated in a single sentence
- Use text when a statement or written description can clearly and succinctly convey information and results without the need for tables or graphs
- When you want to leave no room for misinterpretation by the audience
- Use text when communication is limited to text
  - Abstracts
  - Press Release
  - Some social media (e.g. tweets)

When to use text

- Use text for simple or headline results which can be stated in a single sentence
- Use text when a statement or written description can clearly and succinctly convey information and results without the need for tables or graphs
- When you want to leave no room for misinterpretation by the audience
- Use text when communication is limited to text
  - Abstracts
  - Press Release
  - Some social media (e.g. tweets)

- Examples

  - ***"44% response rate observed in patients receiving 25 mg Active Investigational Product; of these responders all experienced at least 60% tumor shrinkage within 16 weeks"***

  - ***"Grade 3-4 treatment-related adverse events occurred in 46% of patients on the Active Investigational product and 23% of patients on the Placebo"***

When to use tables

- Tables present data in rows and columns organized in some fashion
- Tables can show raw (collected) data or summarized values
- Compared to figures:
  - The numerical values are more easily located, read and extracted
  - Allow for immediate and accurate computation (e.g., differences and ratios)
  - Less room for misinterpretation

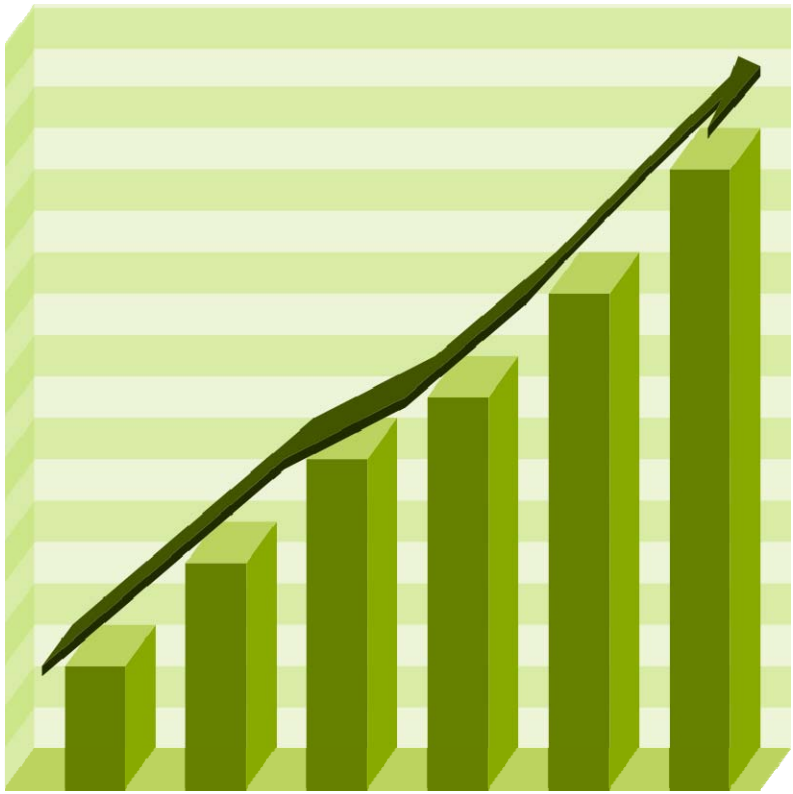
When to use tables

– Example – No need for figure.

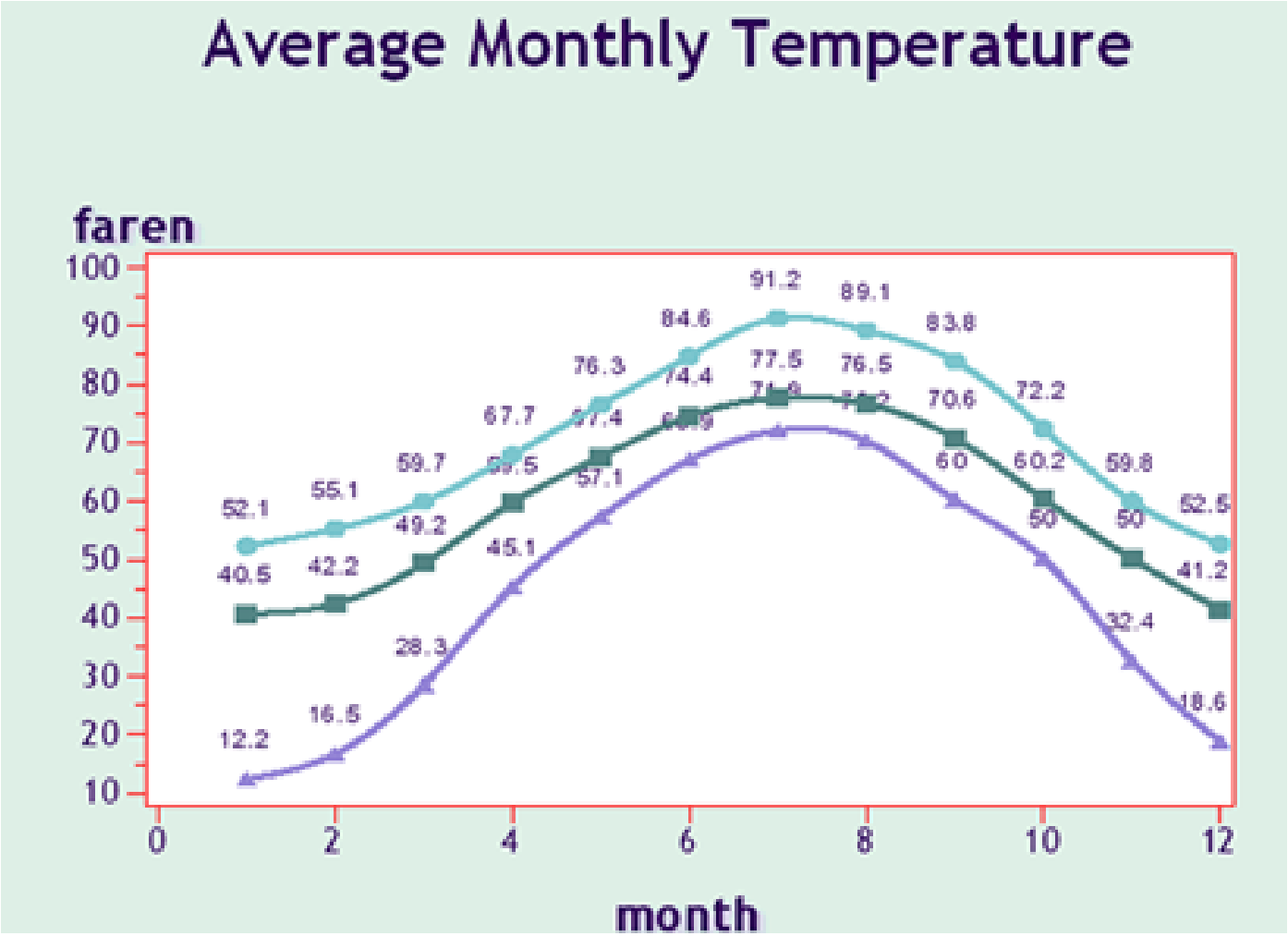
Parameter	Units	Treat 1 n=7	Treat 2 n=9	Treat 3 n=11	Treat 4 n=12
C_{max}	µg/mL	0.034 ± 0.020	0.098 ± 0.046	0.845 ± 0.245	5.271 ± 0.899
T_{max}	Hr	45.2 ± 4.5	65.2 ± 1.9	89.3 ± 8.5	109.1 ± 10.3
AUC_{last}	µg·hr/mL	1.8 ± 1.4	56.8 ± 21.4	345 ± 122	1987 ± 288
$AUC_{0\text{-}inf}$	µg·hr/mL	1.8 ± 1.4	60.2 ± 18.9	346 ± 123	2014 ± 267
CL/F	mL/hr/kg	ND	1.2 ± 0.9	0.8 ± 0.3	0.5 ± 0.2
Vz/F	mL/kg	ND	110 ± 32.4	55.2 ± 20.1	50.0 ± 18.2

When to use figures

- Show patterns, similarities and differences in the data that are not easily picked up by viewing text or tables
  - Present spatially related information (e.g., map)
  - Identify associations, trends, relationships, deviations, minima, maxima and orders of magnitude
  - Hard to identify magnitude of differences from numbers alone
- Take advantage of symbols, lines, colors, sizes, etc to convey related information
- Visualize the by-subject detail to appreciate the granularity and complexity inherent in the data



When to use figures

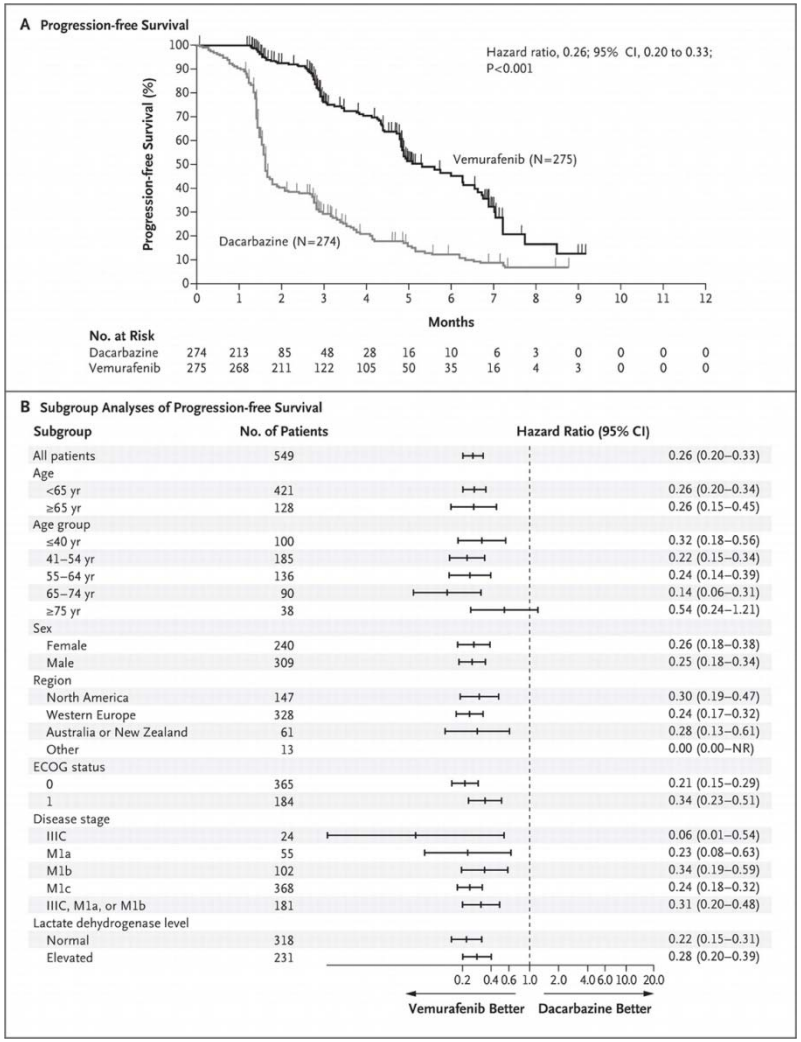


When to use text, tables or figures

- Only include a table or figure if it can more succinctly supply essential information than can a written description
- Do not include identical information in a table and a graph; and don't repeat the information from a table or graph in the text.
- Several iterations and drafts of the displays are necessary for high quality results
- Peer review from several functions can provide valuable feed back on your displays

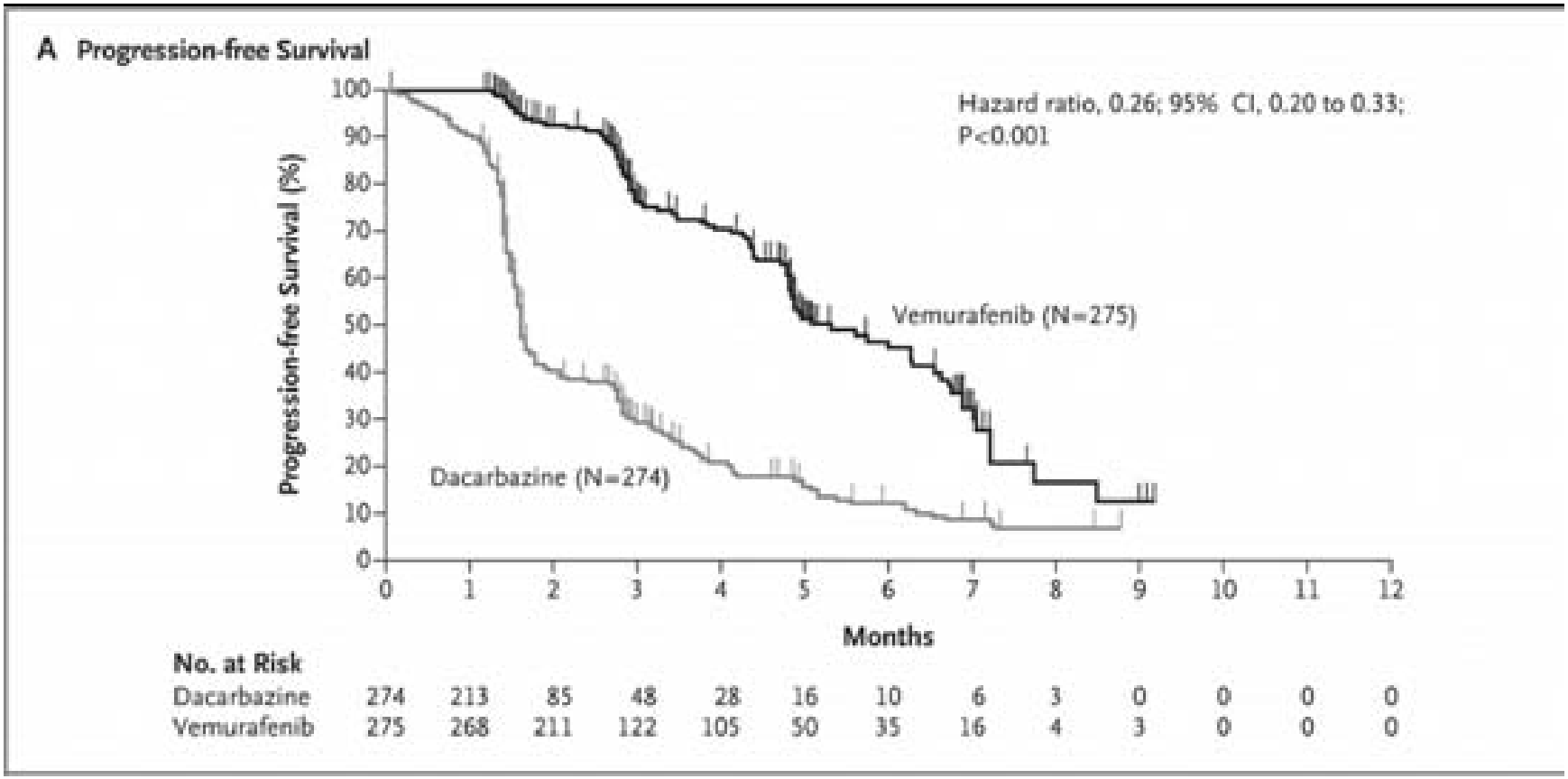
Combining Text, Tables and Figures

Combine Tables and Figures



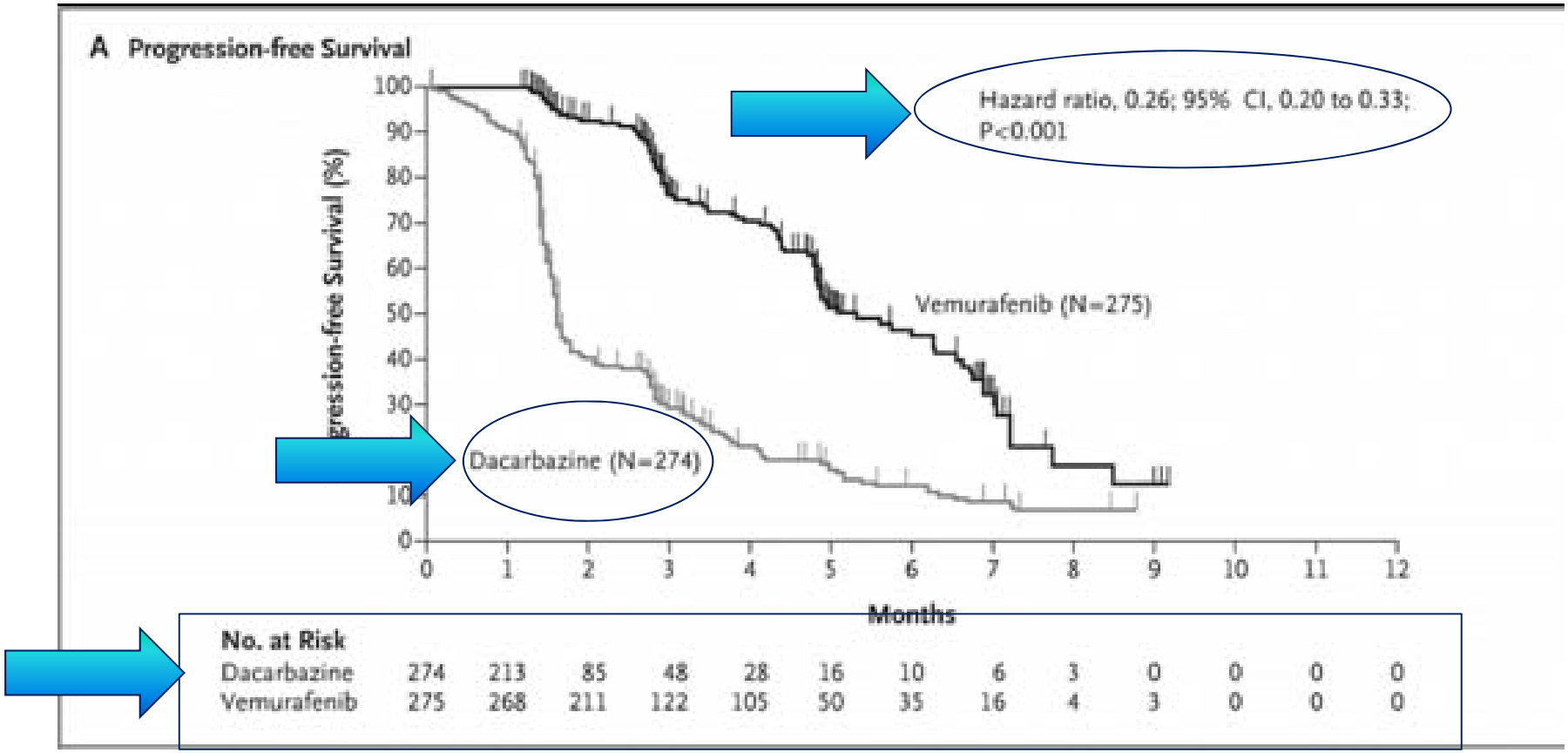
Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Combine Tables and Figures



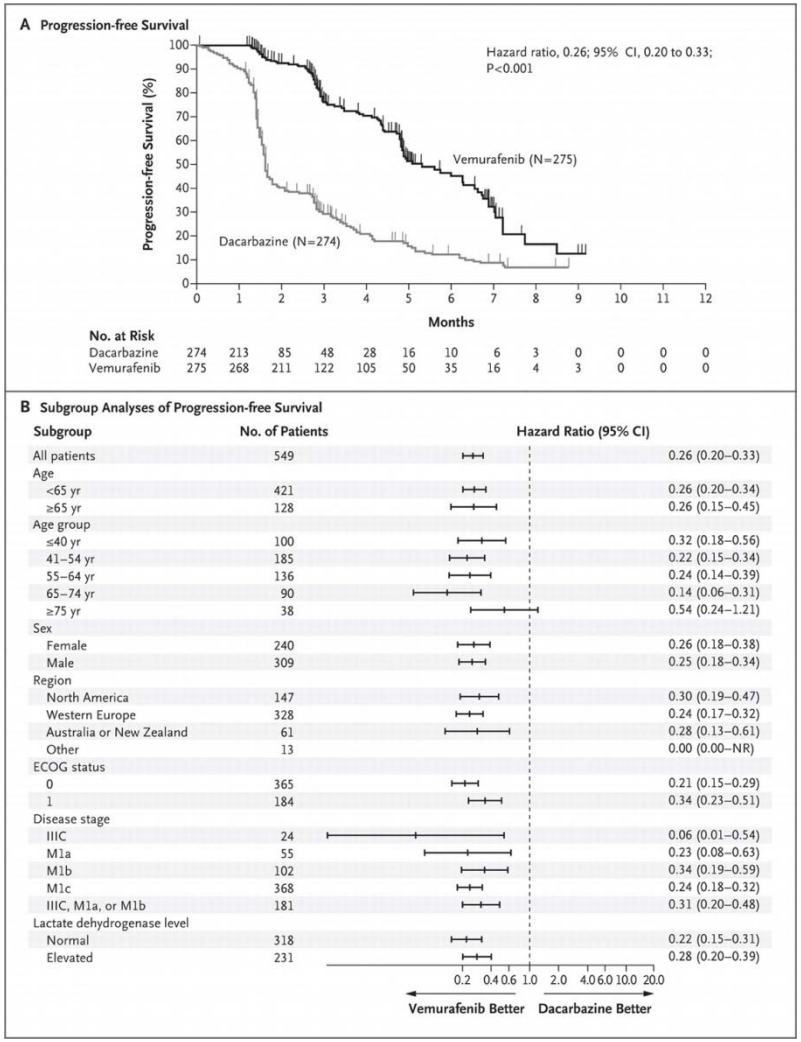
Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Combine Tables and Figures



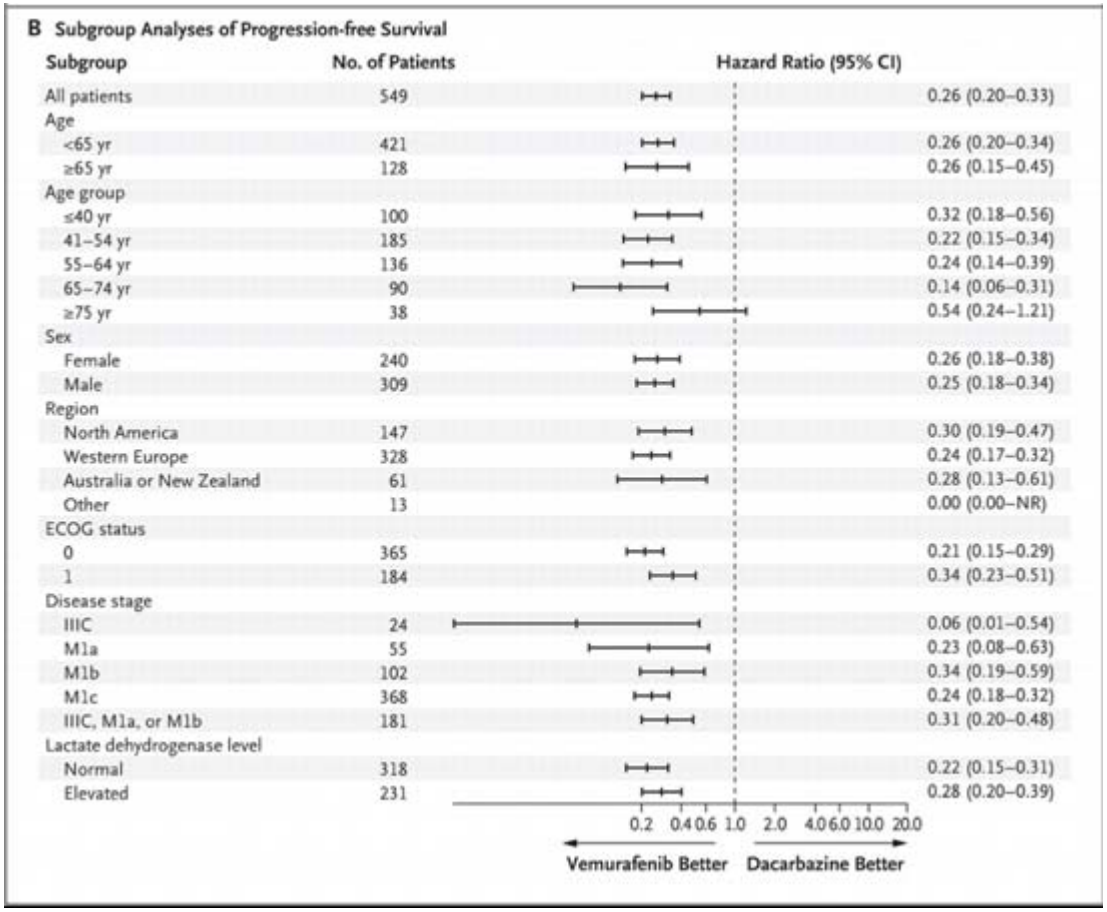
Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Combine Tables and Figures



Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Combine Tables and Figures

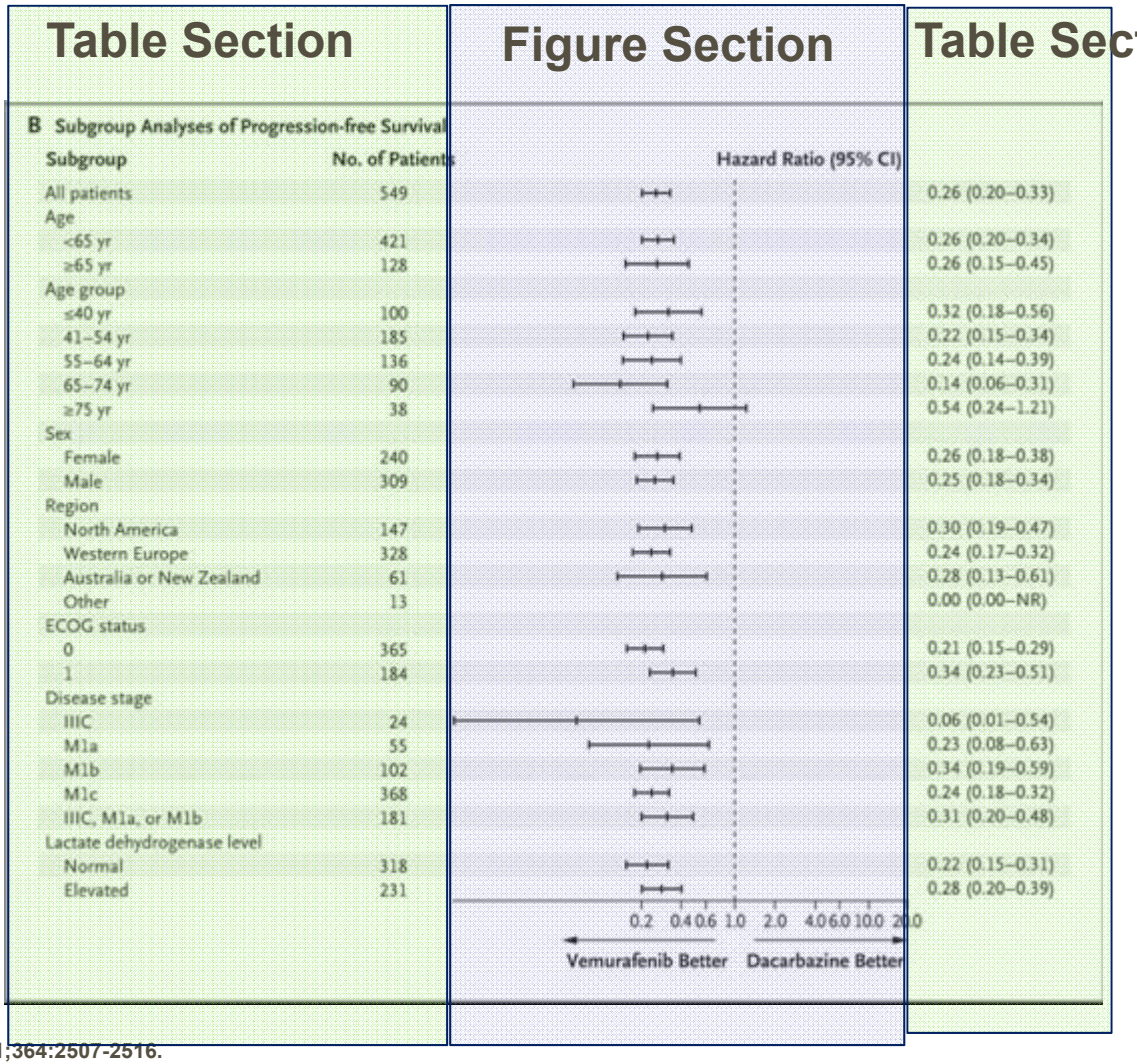


B Subgroup Analyses of Progression-free Survival

Subgroup	No. of Patients	Hazard Ratio (95% CI)
All patients	549	0.26 (0.20–0.33)
Age		
<65 yr	421	0.26 (0.20–0.34)
≥65 yr	128	0.26 (0.15–0.45)
Age group		
≤40 yr	100	0.32 (0.18–0.56)
41–54 yr	185	0.22 (0.15–0.34)
55–64 yr	136	0.24 (0.14–0.39)
65–74 yr	90	0.14 (0.06–0.31)
≥75 yr	38	0.54 (0.24–1.21)
Sex		
Female	240	0.26 (0.18–0.38)
Male	309	0.25 (0.18–0.34)
Region		
North America	147	0.30 (0.19–0.47)
Western Europe	328	0.24 (0.17–0.32)
Australia or New Zealand	61	0.28 (0.13–0.61)
Other	13	0.00 (0.00–NR)
ECOG status		
0	365	0.21 (0.15–0.29)
1	184	0.34 (0.23–0.51)
Disease stage		
IIIC	24	0.06 (0.01–0.54)
M1a	55	0.23 (0.08–0.63)
M1b	102	0.34 (0.19–0.59)
M1c	368	0.24 (0.18–0.32)
IIIC, M1a, or M1b	181	0.31 (0.20–0.48)
Lactate dehydrogenase level		
Normal	318	0.22 (0.15–0.31)
Elevated	231	0.28 (0.20–0.39)

Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Combine Tables and Figures

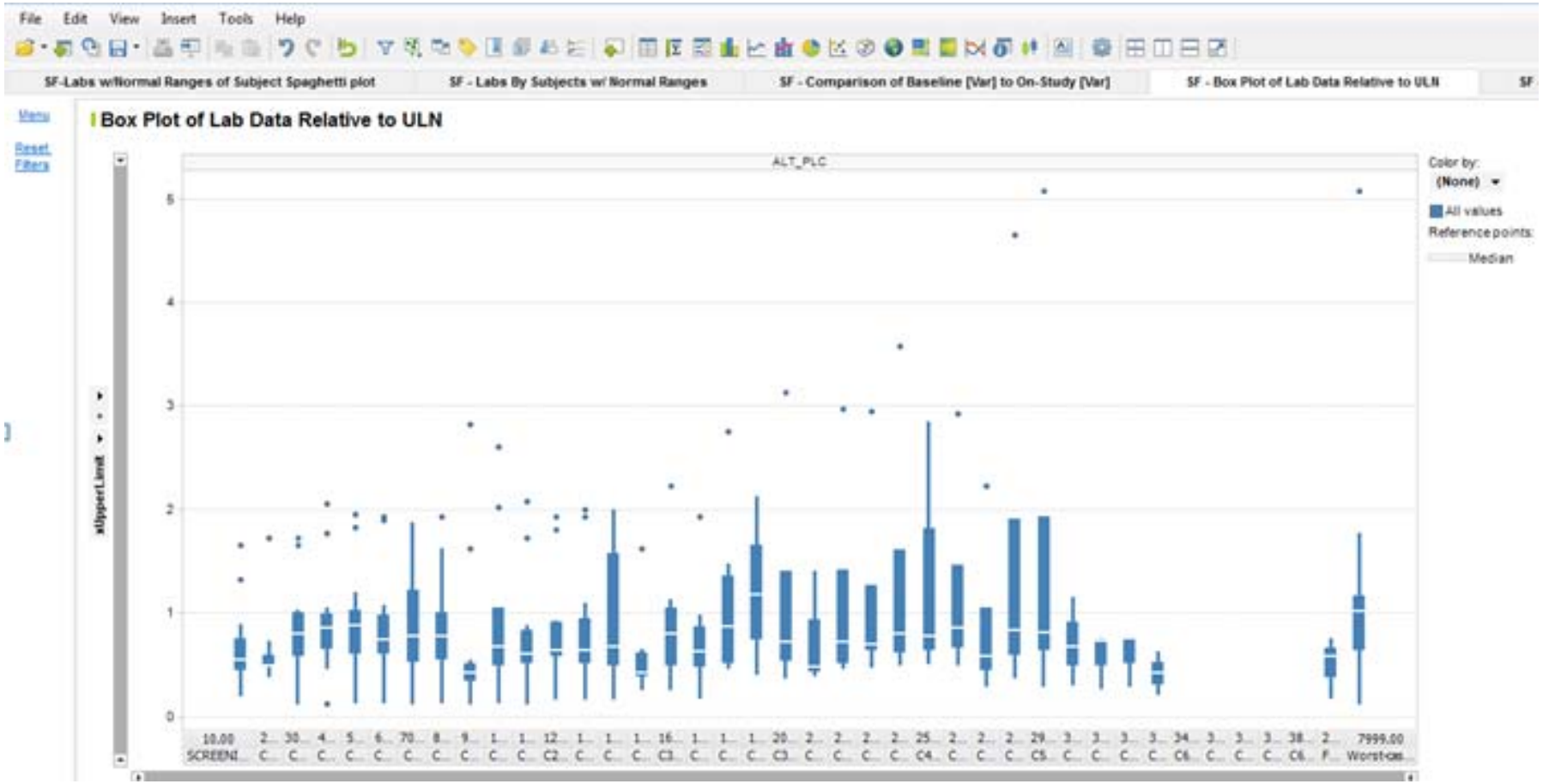


Chapman PB et al. N Engl J Med 2011;364:2507-2516.

Interactive Displays

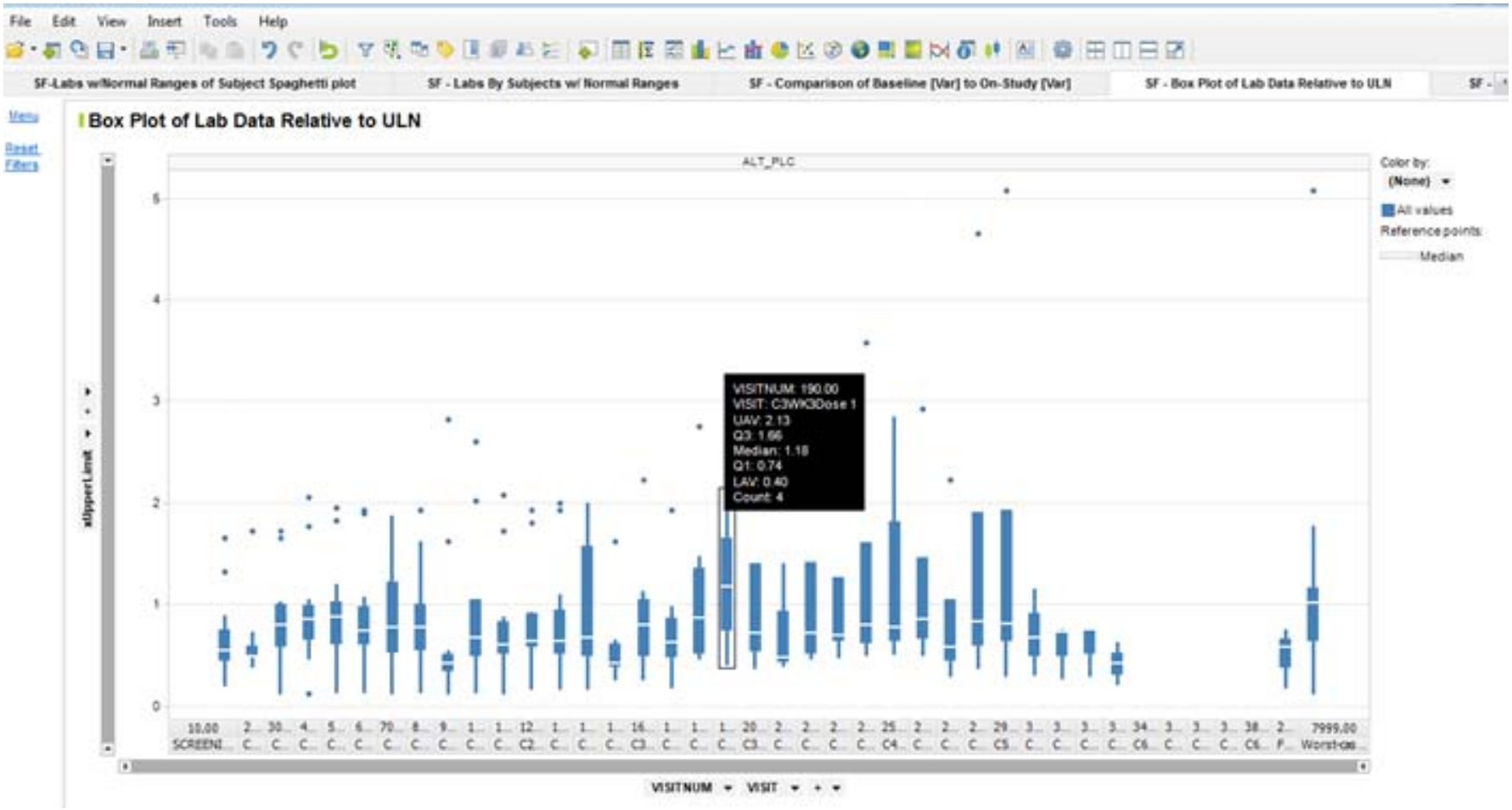
Interactive Figures and Tables

TIBCO® Spotfire



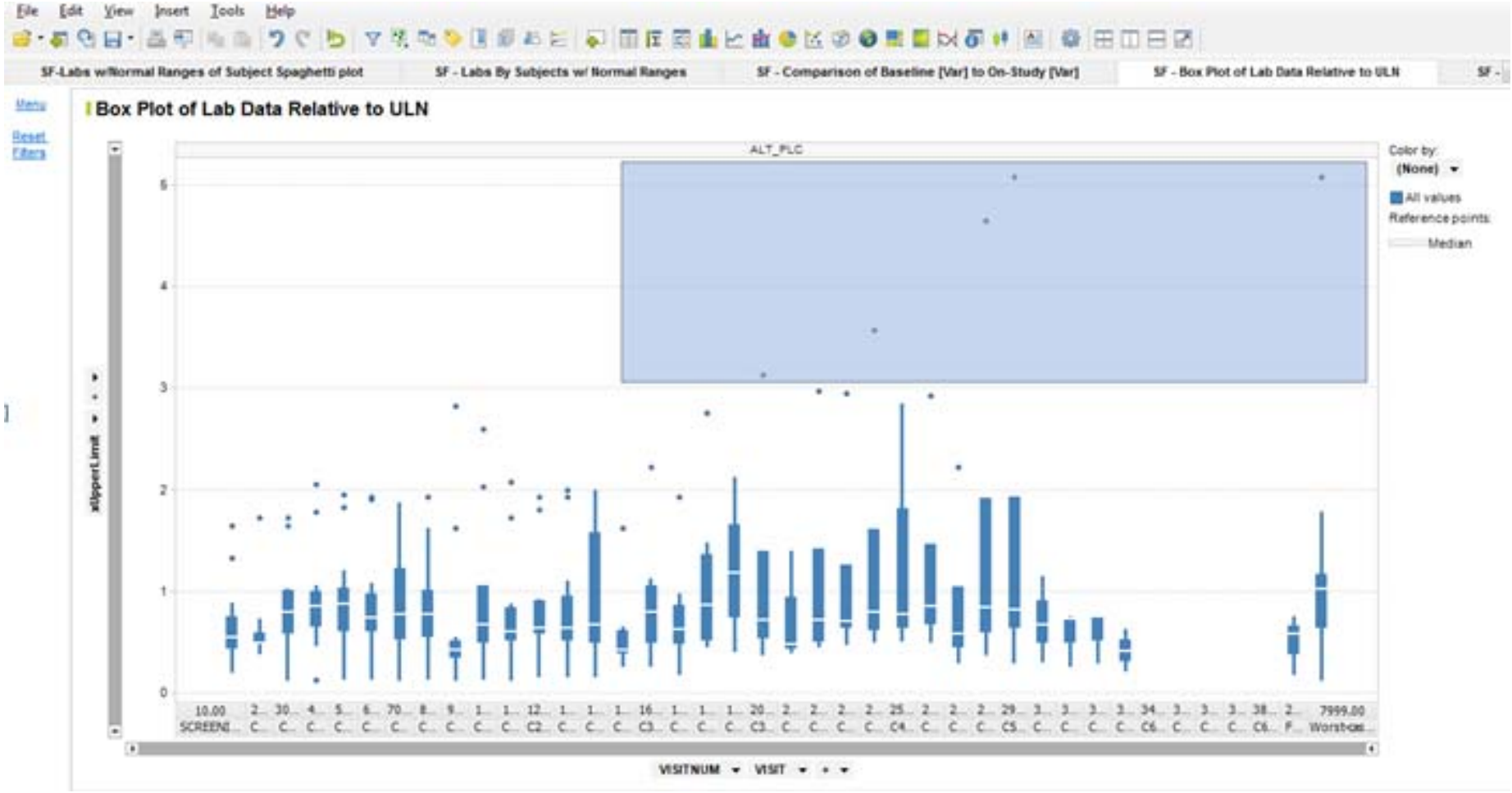
Interactive Figures and Tables

TIBCO® Spotfire



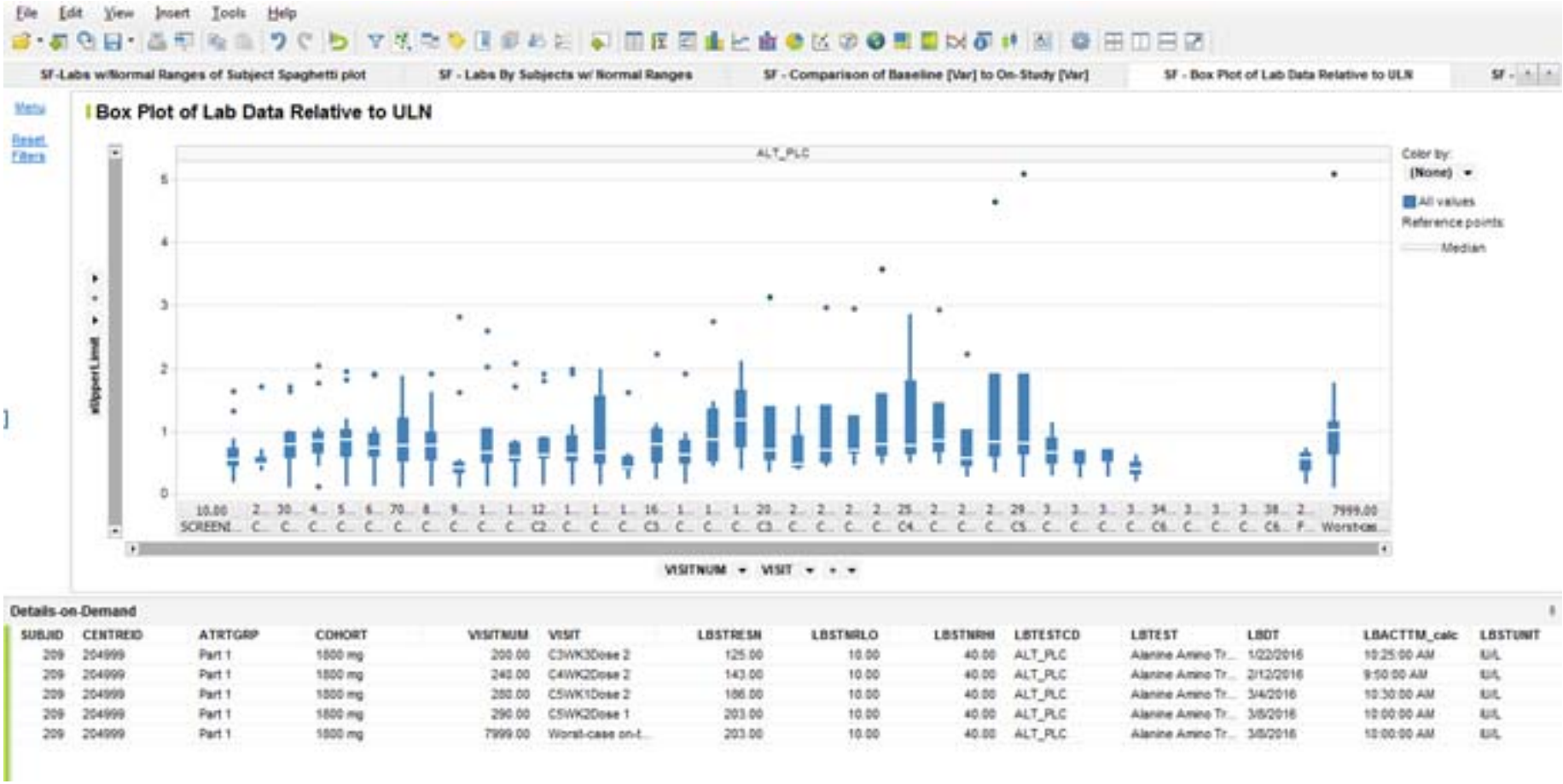
Interactive Figures and Tables

TIBCO® Spotfire



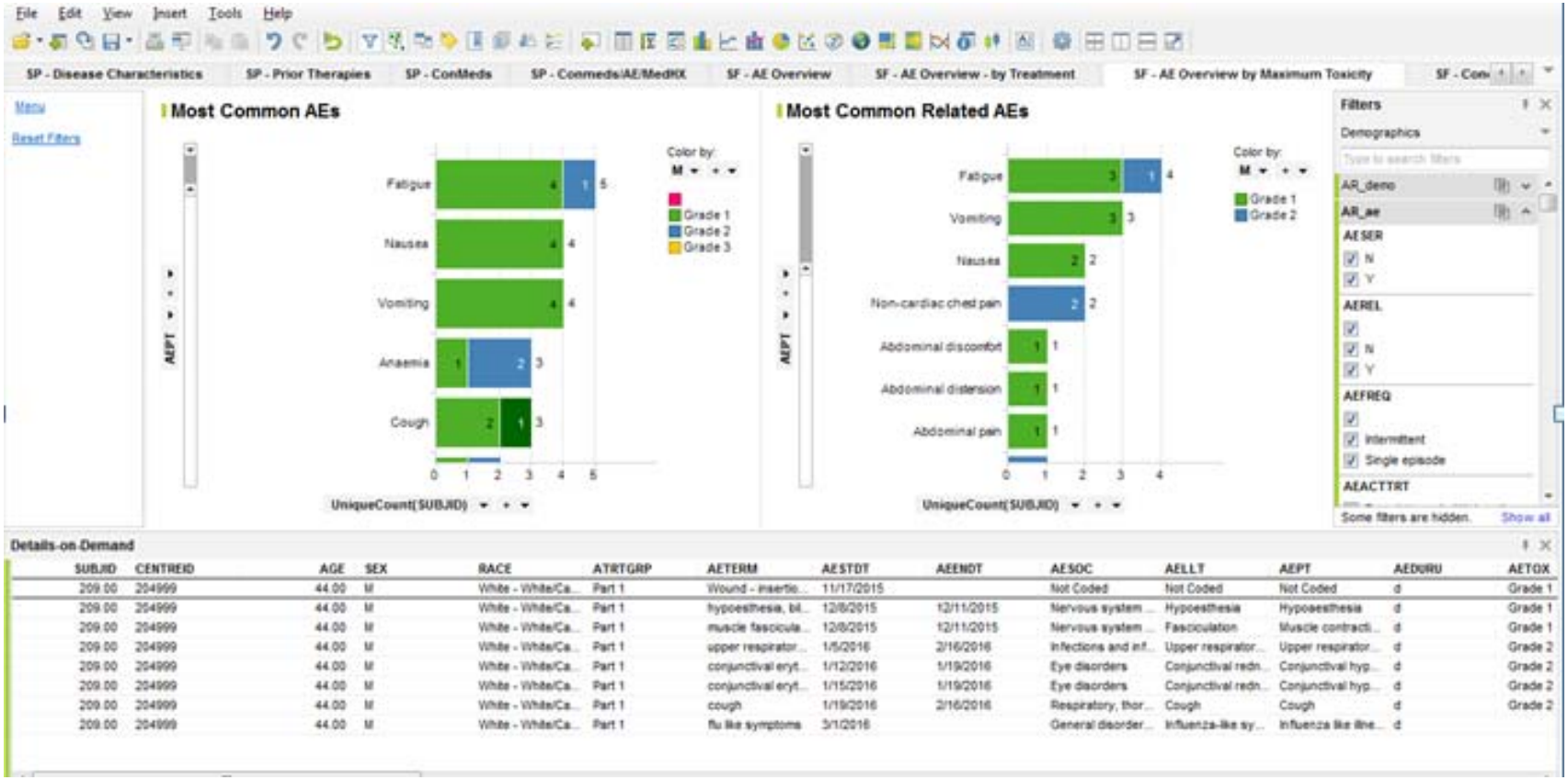
Interactive Figures and Tables

TIBCO® Spotfire



Interactive Figures and Tables

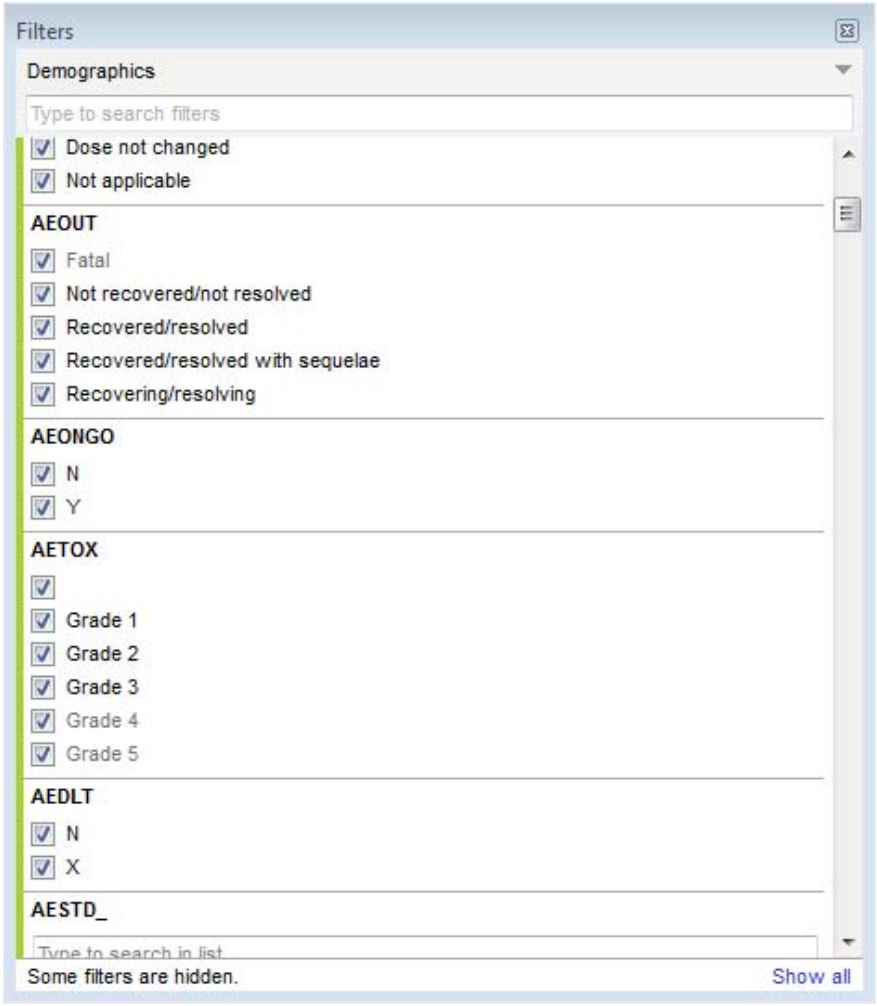
TIBCO® Spotfire



Interactive Figures and Tables

TIBCO® Spotfire

- Interactivity can have costs
  - Interactive displays put the burden of display choice and design on the user
  - Time and attention needed to decide to subtract irrelevant information
  - Users many not discover or know all the interactive functions available
  - Ensure controls are in place to prevent unblinding/unintended bias
  - Users may forget to reset filters or markings



How to choose the correct figure

Factors in determining which graph to create

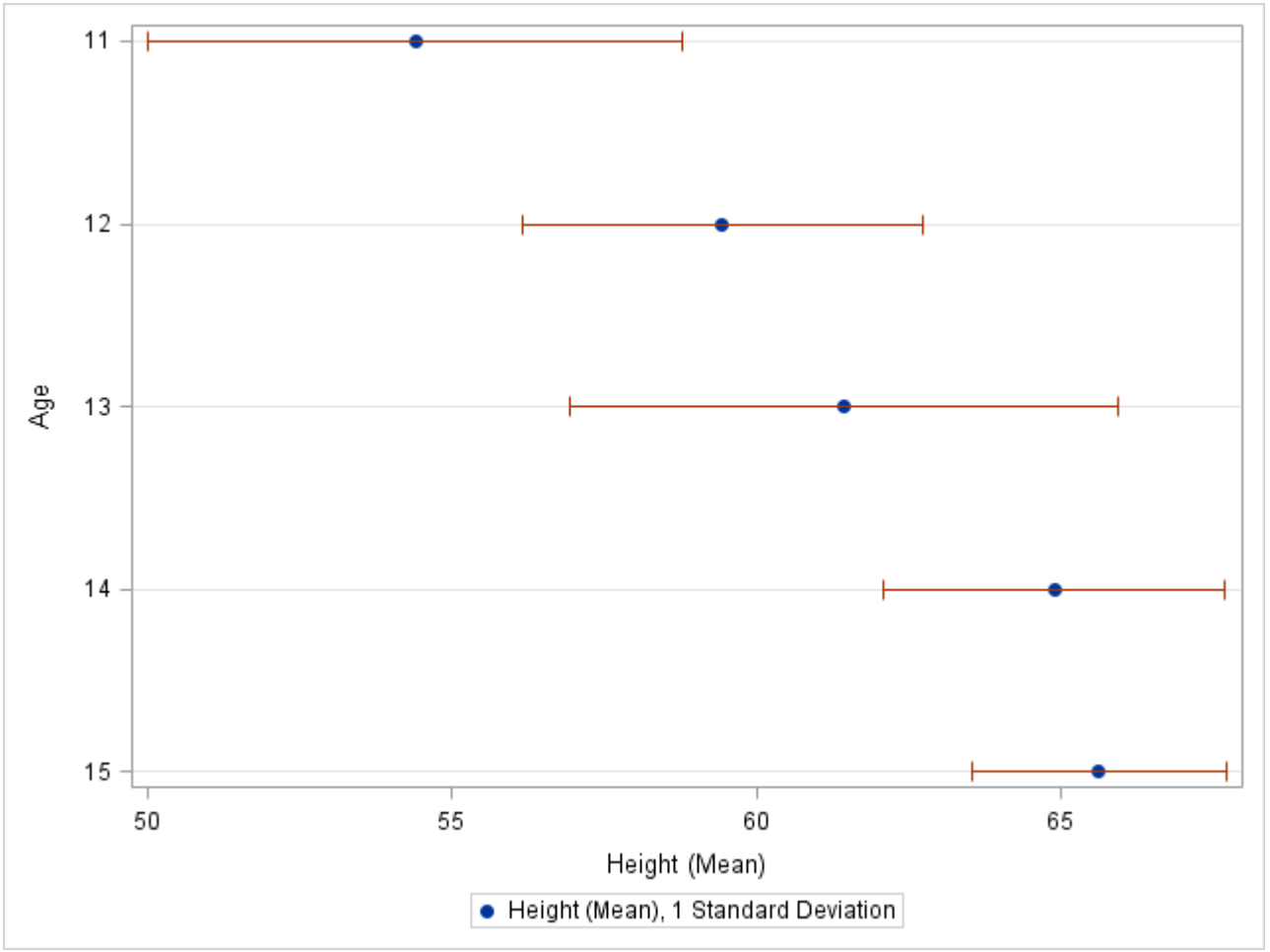
- Audience or decision makers
- How many variables of interest (univariate, bivariate, multivariate)
- Quantitative or Categorical variables
  - Quantitative: What level of detail about the distribution of the data to be displayed
  - Categorical: How many levels of the categorical variables
- Number of grouping variables or subsets to be compared

Quantitative (Continuous) Variables

- Summary Level
  - If the detail of the distribution of the data is not important (e.g., high level summaries of normally distributed data with large number of samples)
  - Low level of detail about distribution of the data
  - High level communication
  - Examples: Bar plots, line plots, dot plots and forest plots

Quantitative (Continuous) Variables

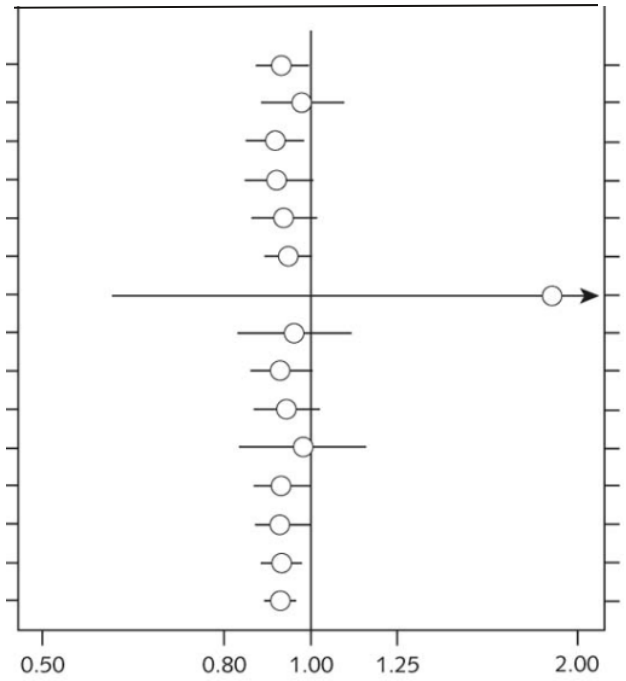
Graphs with low level of detail about the distribution of the data

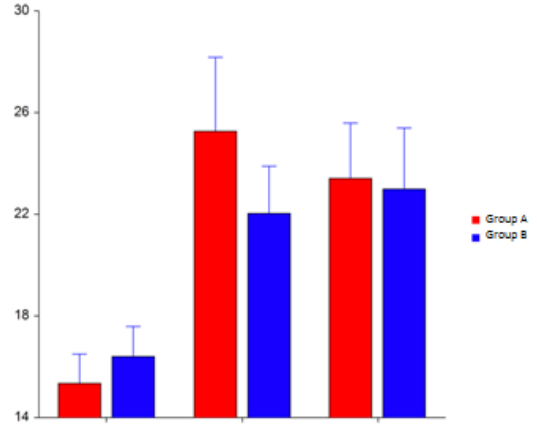


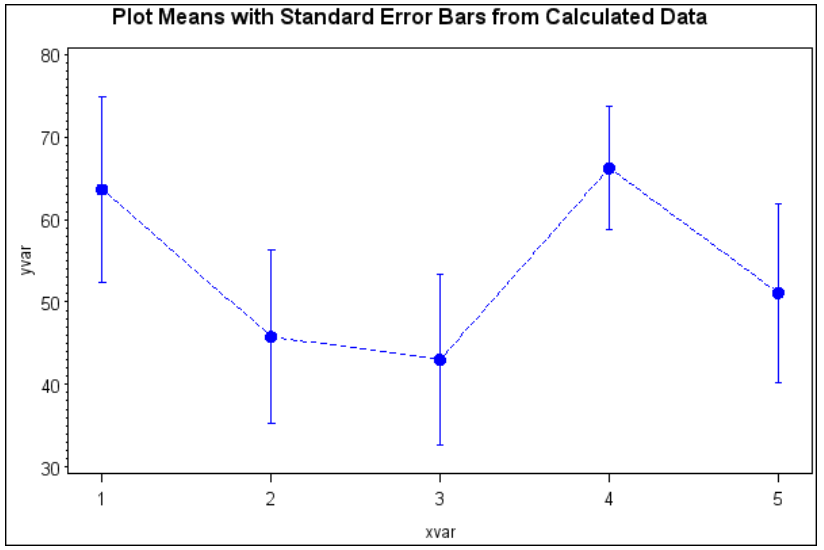
Quantitative (Continuous) Variables

Graphs with low level of detail about the distribution of the data

- Forest Plot
- Bar Plot
- Line Plot





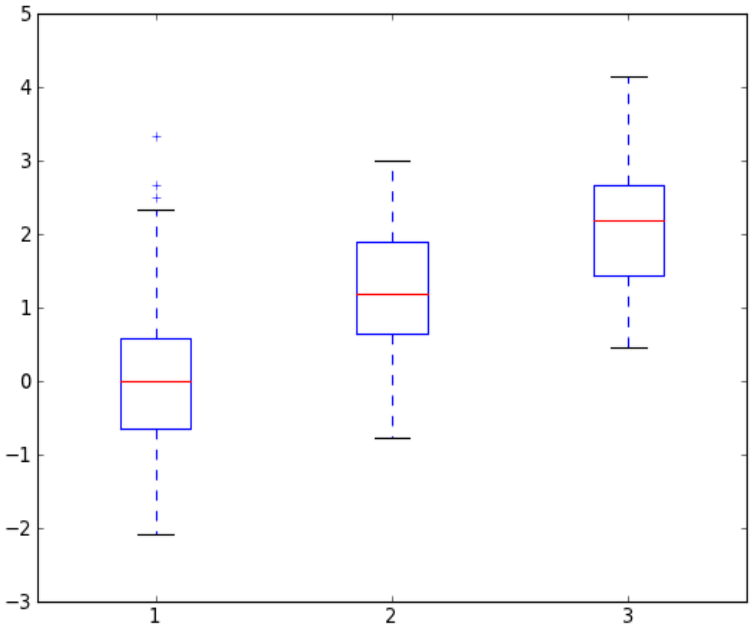


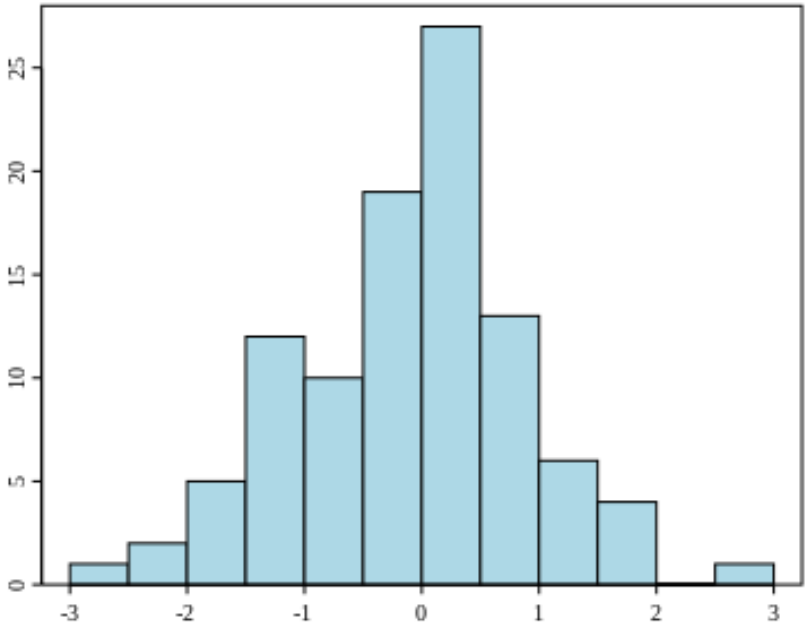
Quantitative (Continuous) Variables

- Medium level of detail about your data's distribution
  - More detail than summary level but not as much other graph types.
  - Examples: Box plots and histograms

Quantitative (Continuous) Variables

- Medium level of detail about your data's distribution
  - More detail than summary level but not as much other graph types.
  - Examples: Box plots and histograms



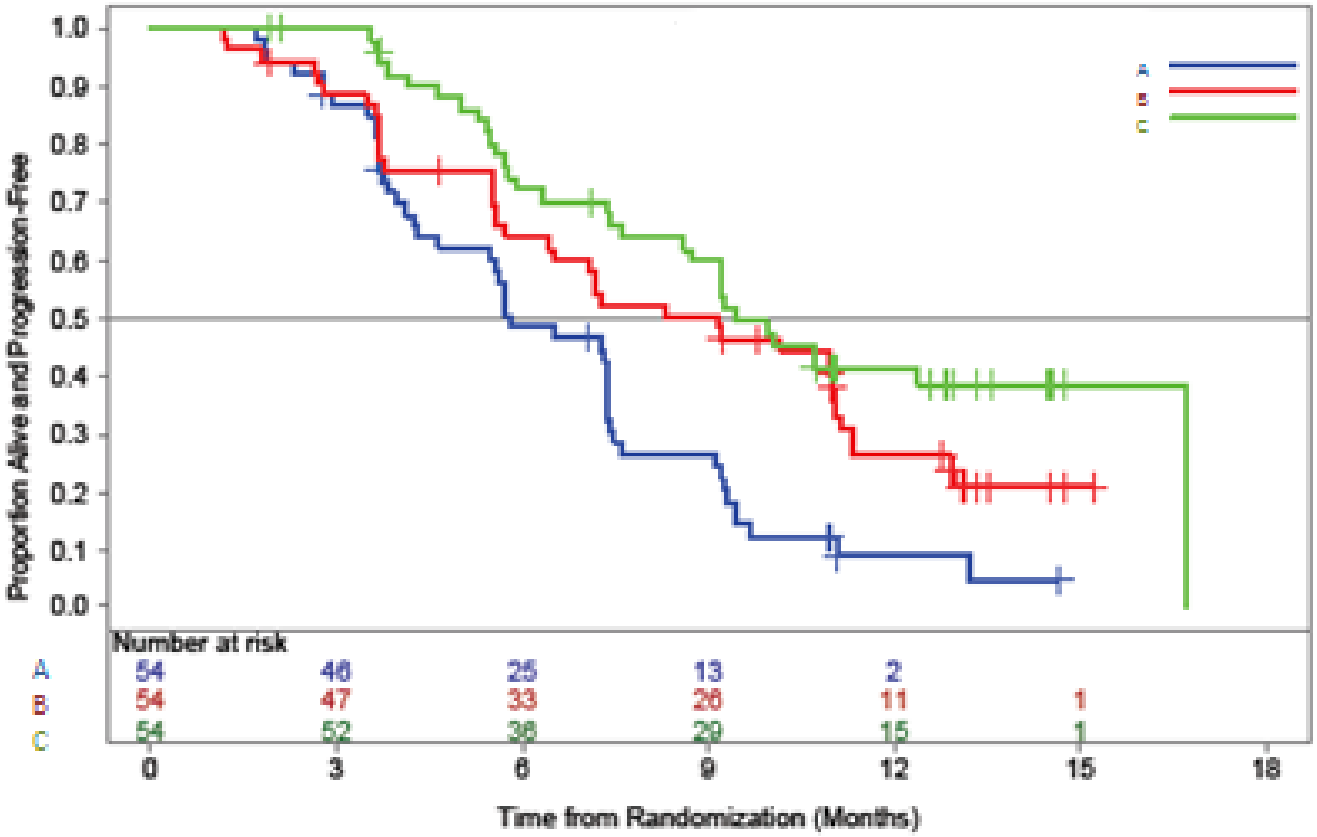


Quantitative (Continuous) Variables

- Detail Level
  - If the detail of the distribution of the data is very important (fine grained detail)
  - Raw data
  - Analysis and exploration
  - Examples: KM plot, cumulative distribution plot, violin plot

Quantitative (Continuous) Variables

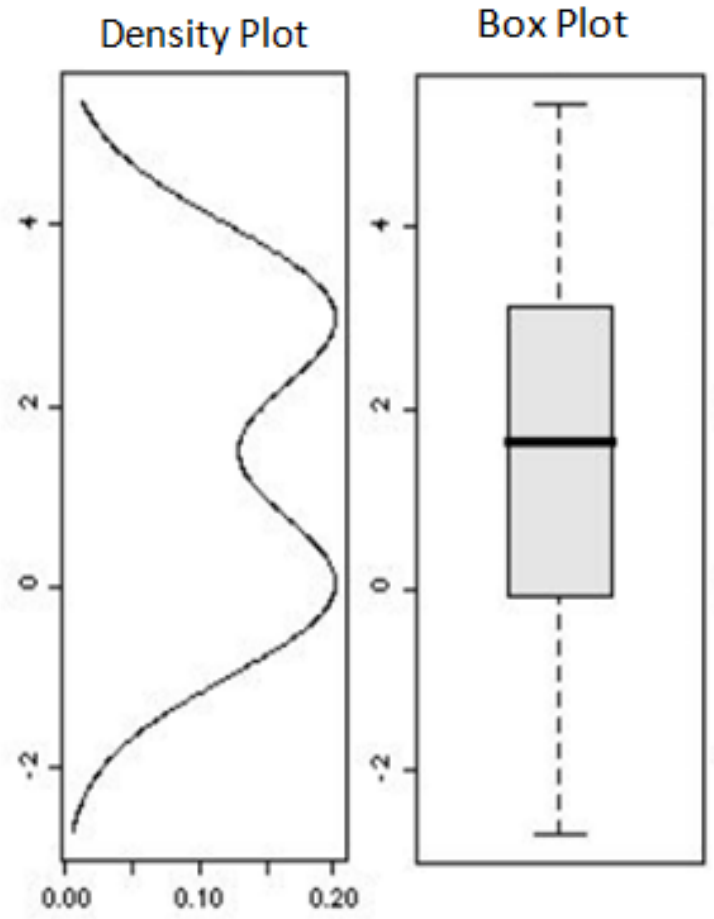
Detail Level – KM Plot



Quantitative (Continuous) Variables

High level of detail about distribution – Violin Plots

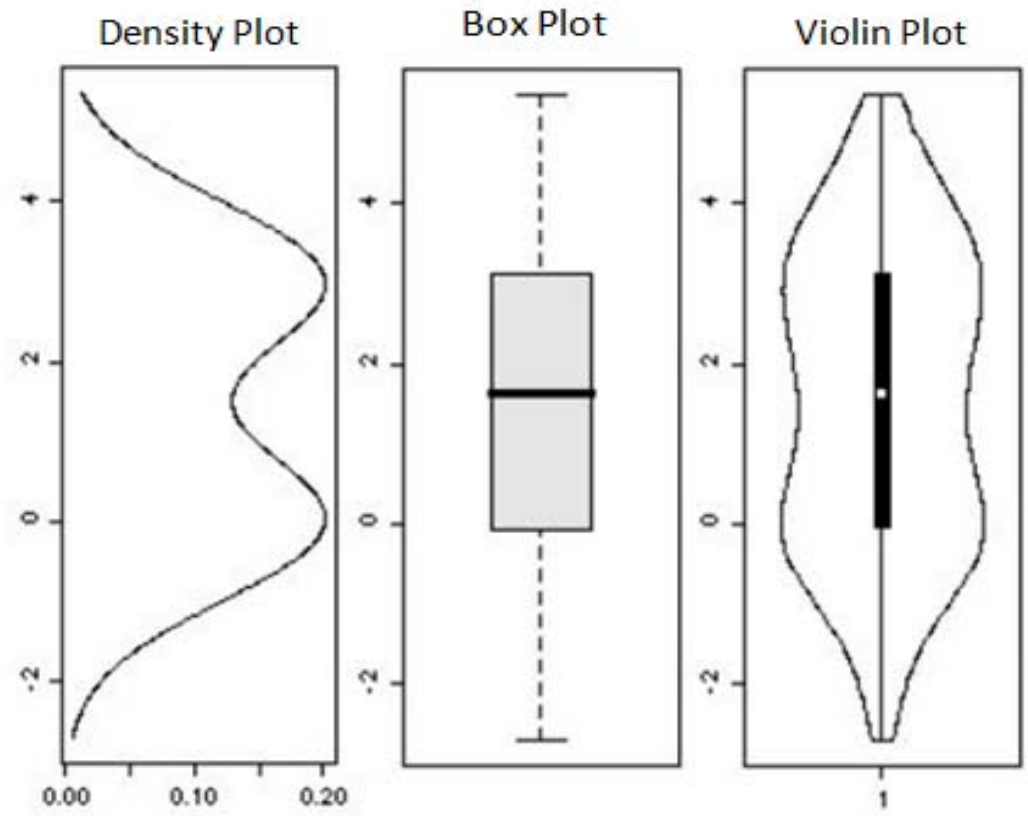
– Box plot with a rotated kernel density plot on each side



Quantitative (Continuous) Variables

High level of detail about distribution – Violin Plots

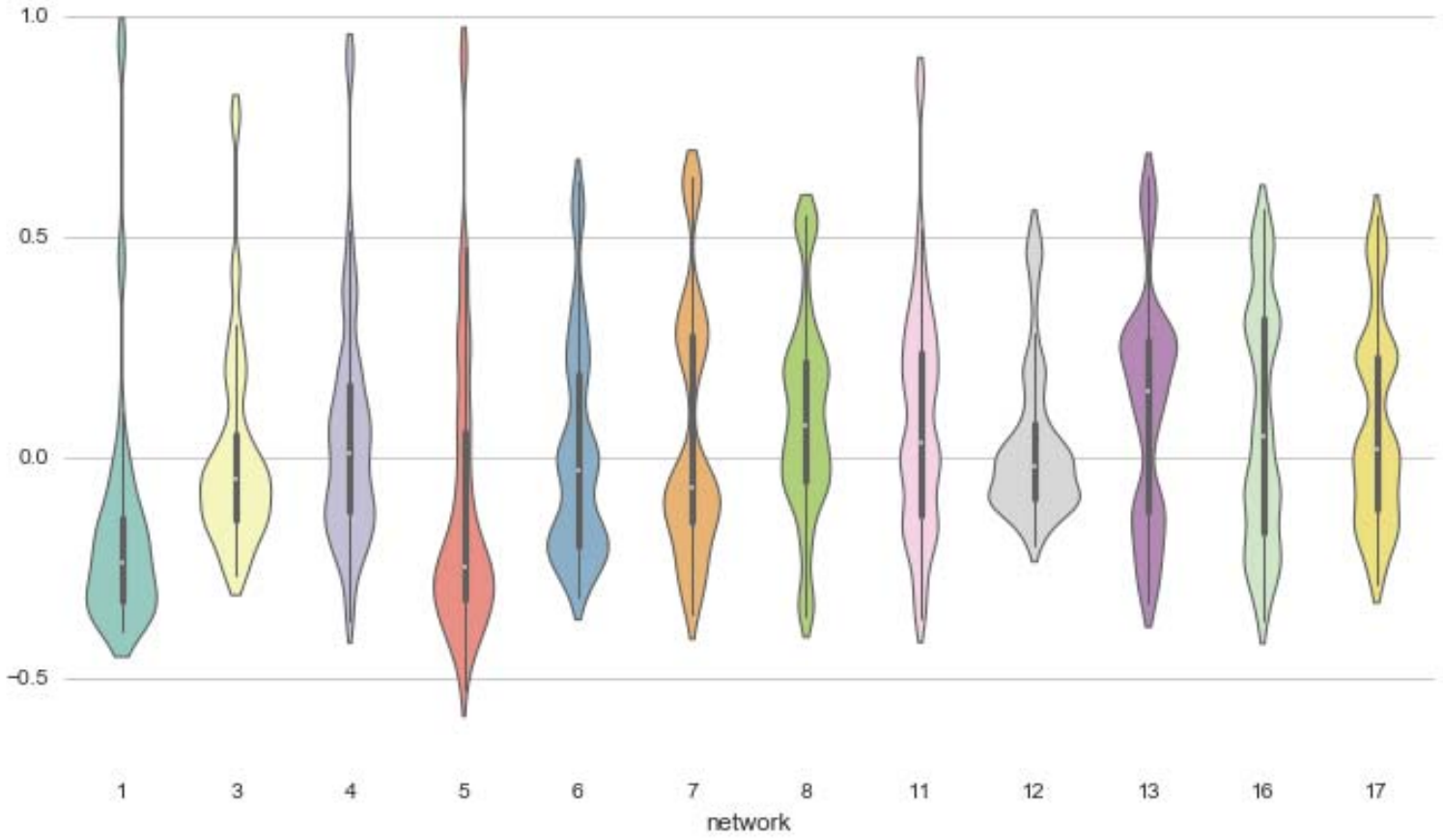
– Box plot with a rotated kernel density plot on each side



Quantitative (Continuous) Variables

High level of detail about distribution – Violin Plots

– Box plot with a rotated kernel density plot on each side



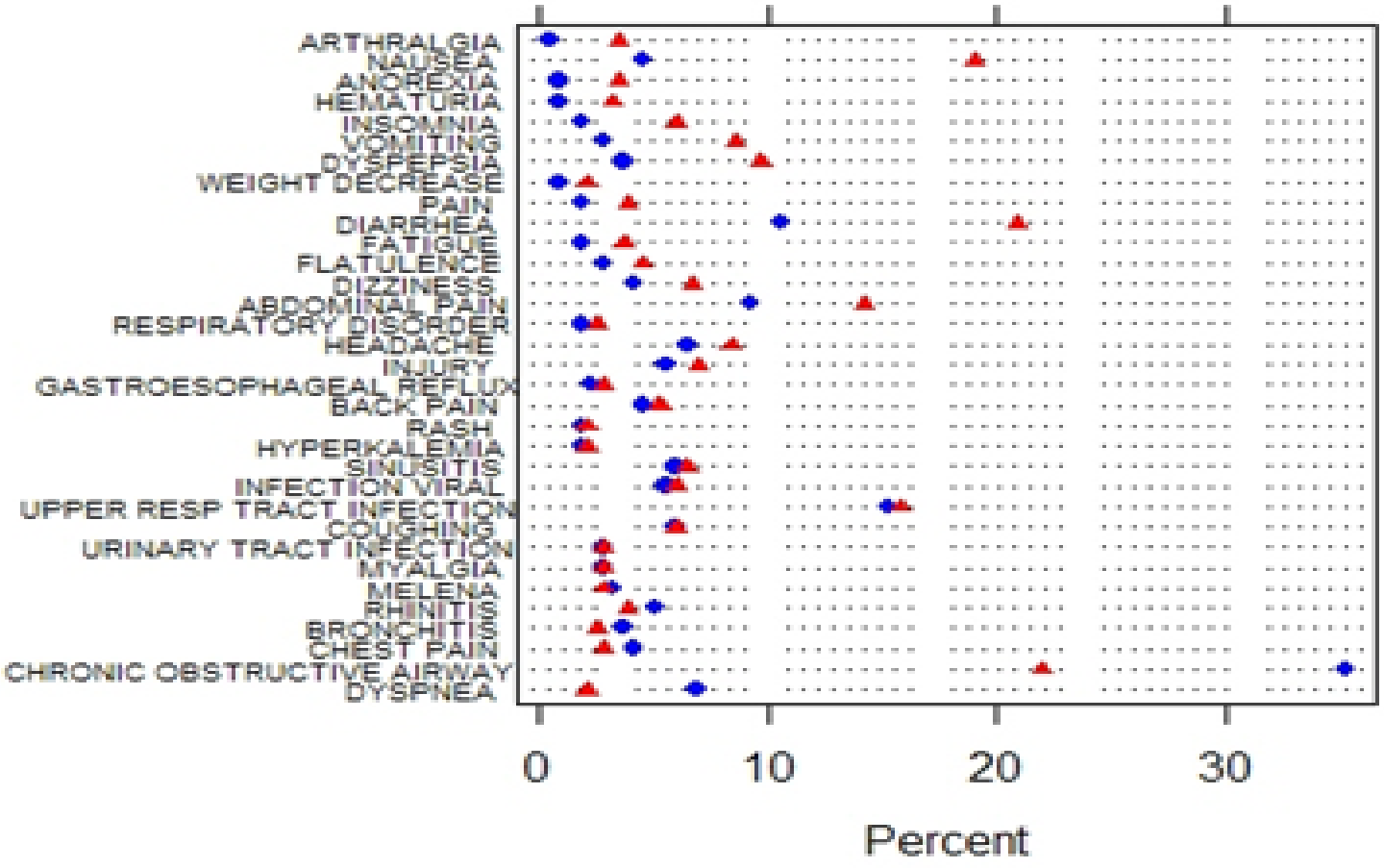
https://stanford.edu/~mwaskom/software/seaborn/examples Accessed 26JUL2016

Categorical variables

- Distribution of categorical variables is less informative than continuous
- Describe the probability of each level of the variable
- Binary Variables (yes/no, on/off, 1/0)
  - Probability of one or the other (e.g., probability of yes or no)
  - Since the total probability adds to 1, one can easily graph the single possibility without need to graph the other
  - Examples: Bar or dot plots
  - Group many binary variables together (multiple variable dot plot AE)

Categorical variables

Binary Variables



Categorical variables

Higher dimension categorical variables (>2 possibilities)

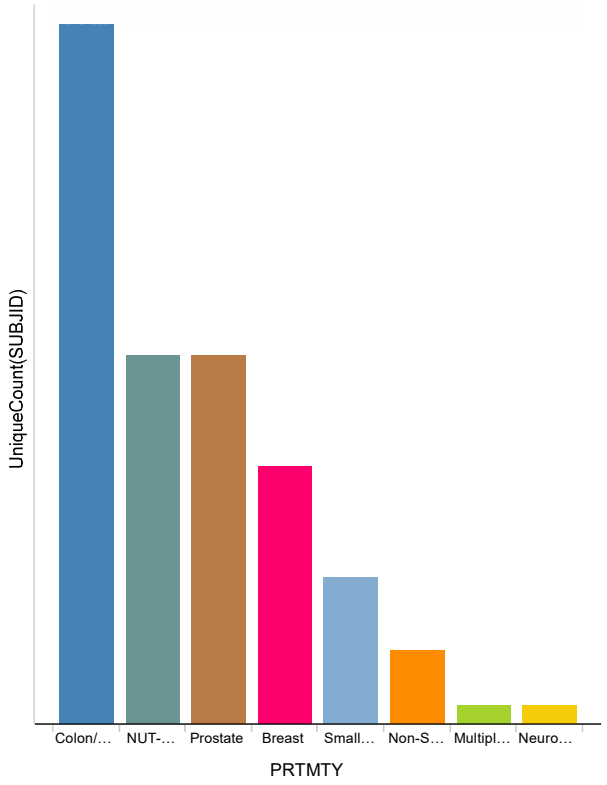
- Probability of each level should be shown
- Examples: Bar, dot and pie plots

Categorical variables

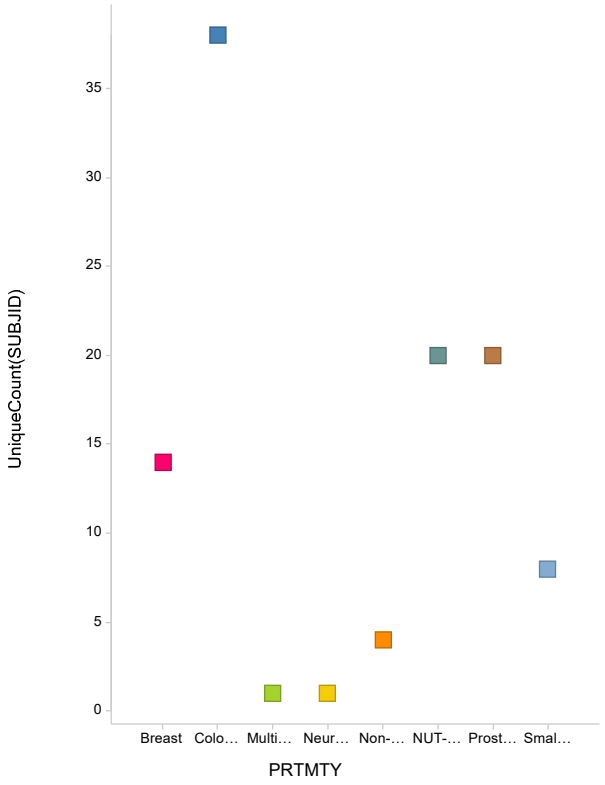
Higher dimension categorical variables (>2 possibilities)

- Probability of each level should be shown
- Examples: Bar, dot and pie plots

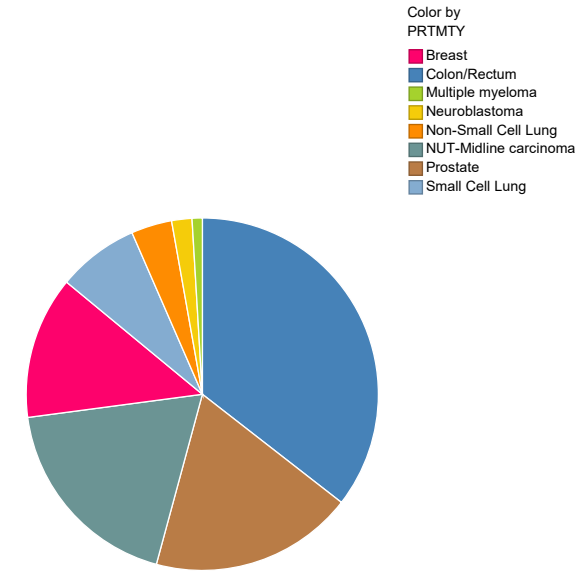
Bar Chart



Dot Plot



Pie Chart

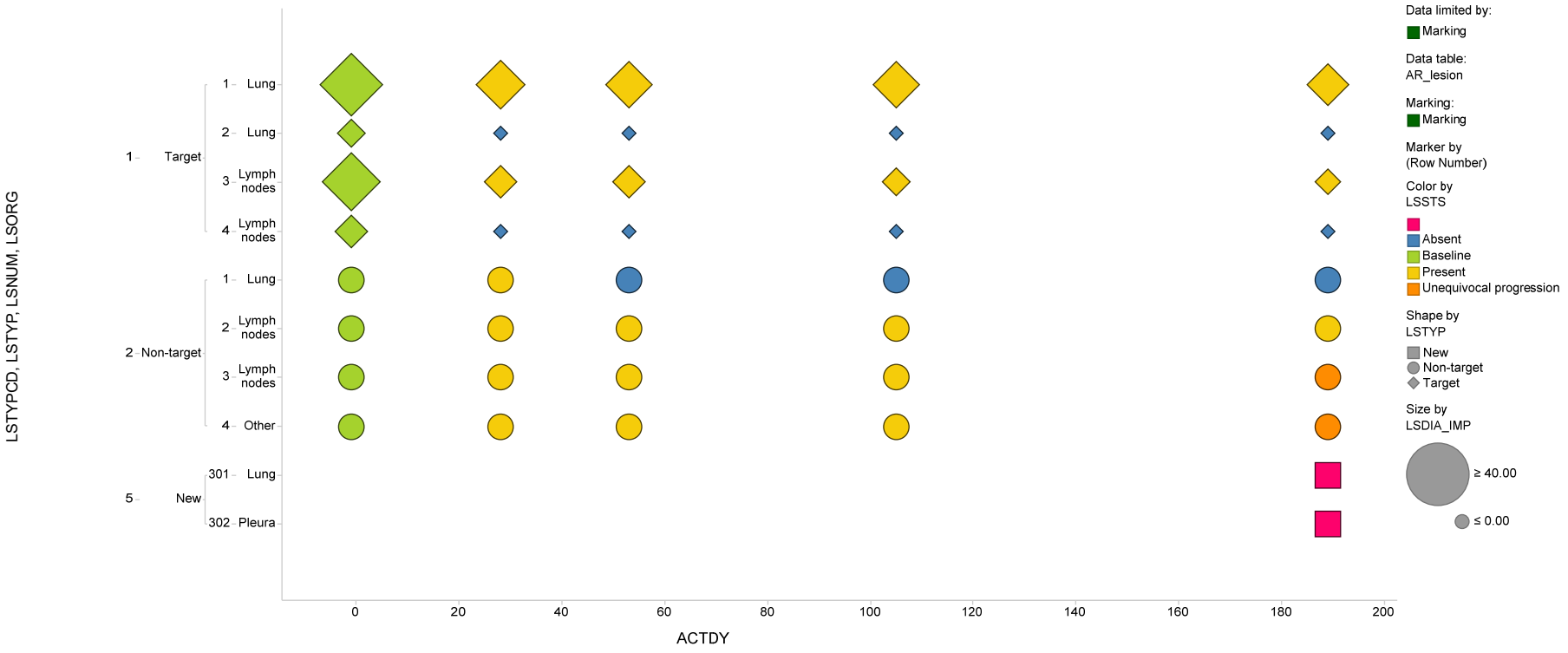


Grouping variables

- Display different levels/groups/subsets
- Number of groups or levels
- Fewer number of groups
  - Vary the visual properties of each group identified in a legend
    - Symbol type
    - Symbol color
    - Error bar color
    - Symbol size
    - Line type

Grouping variables

Use Visual Properties to identify related items



Grouping variables

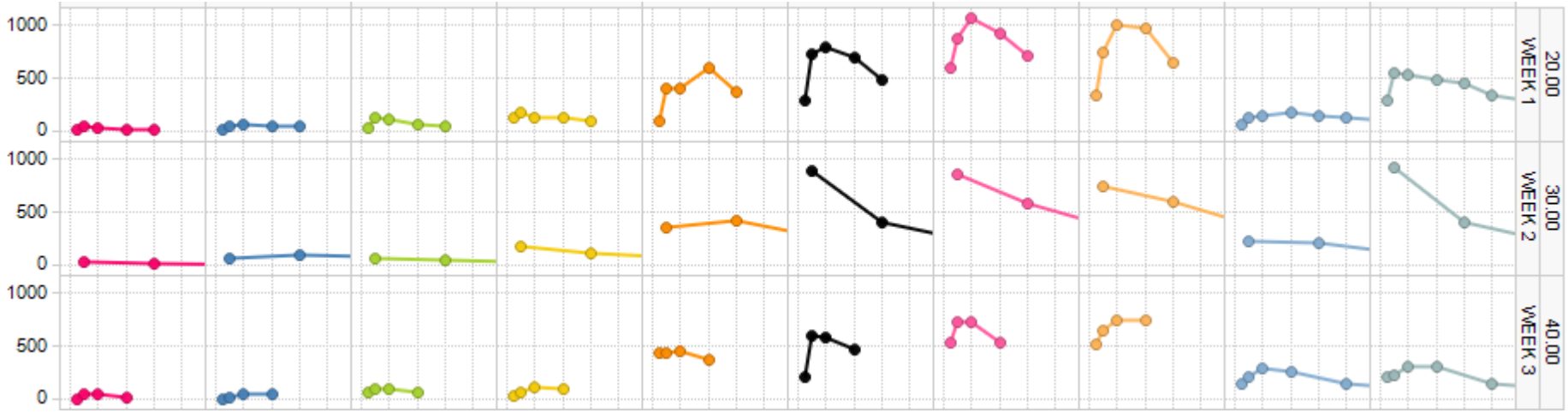
Greater number of groups/levels

– Split the plots out by the group to avoid clutter/complexity
– Trellis or Multi-panel plots

Grouping variables

Greater than 4 groups/levels

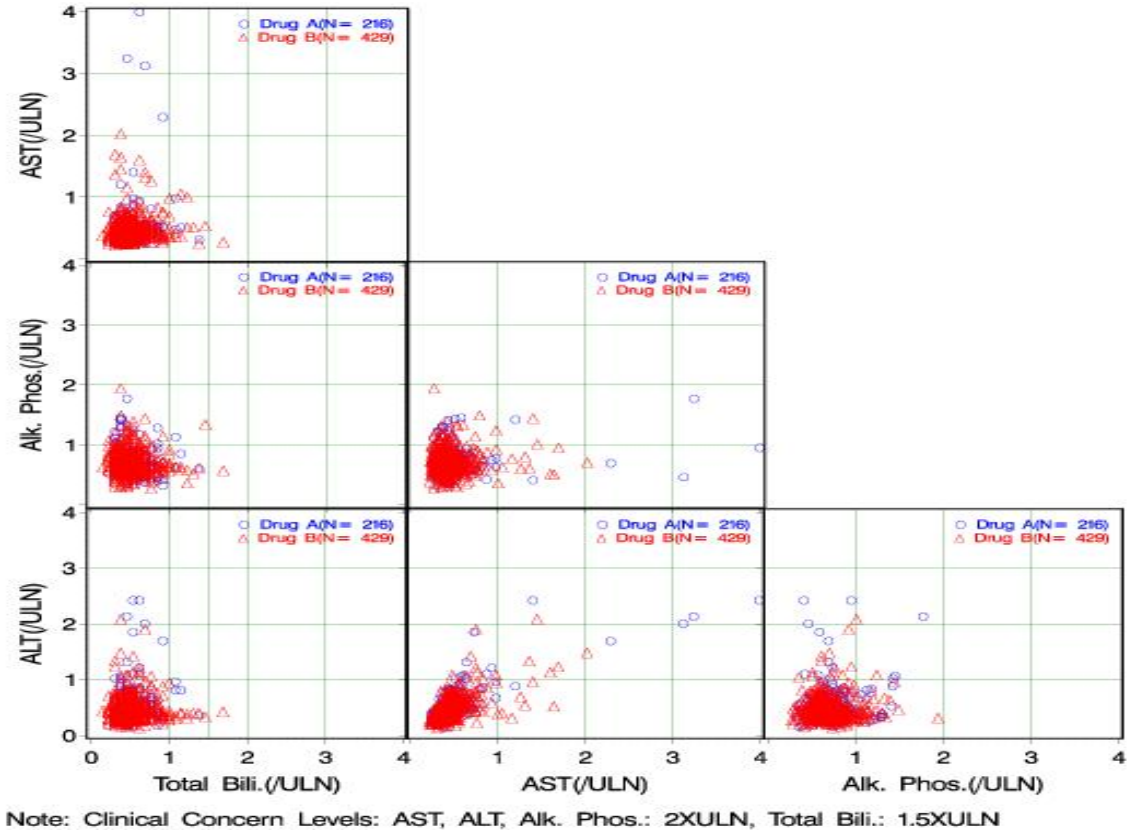
- Split the plots out by the group to avoid clutter/complexity
- Trellis or Multi-panel plots



Grouping variables

Greater than 4 groups/levels

- Split the plots out by the group to avoid clutter/complexity
- Trellis or Multi-panel plots



Grouping variables

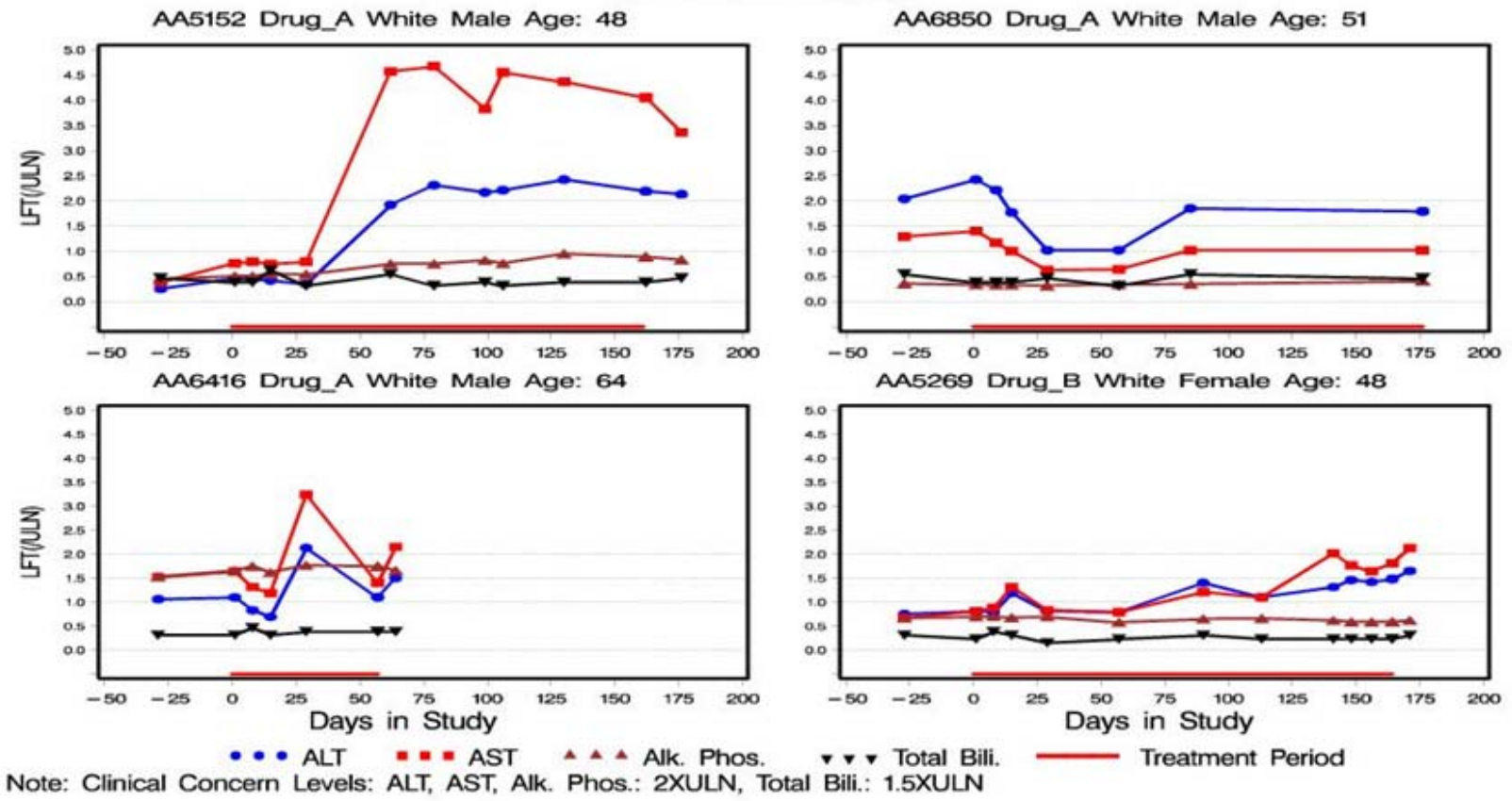
As the number of levels increase combine all these elements

Grouped dot plots, grouped KM plots

Grouping variables

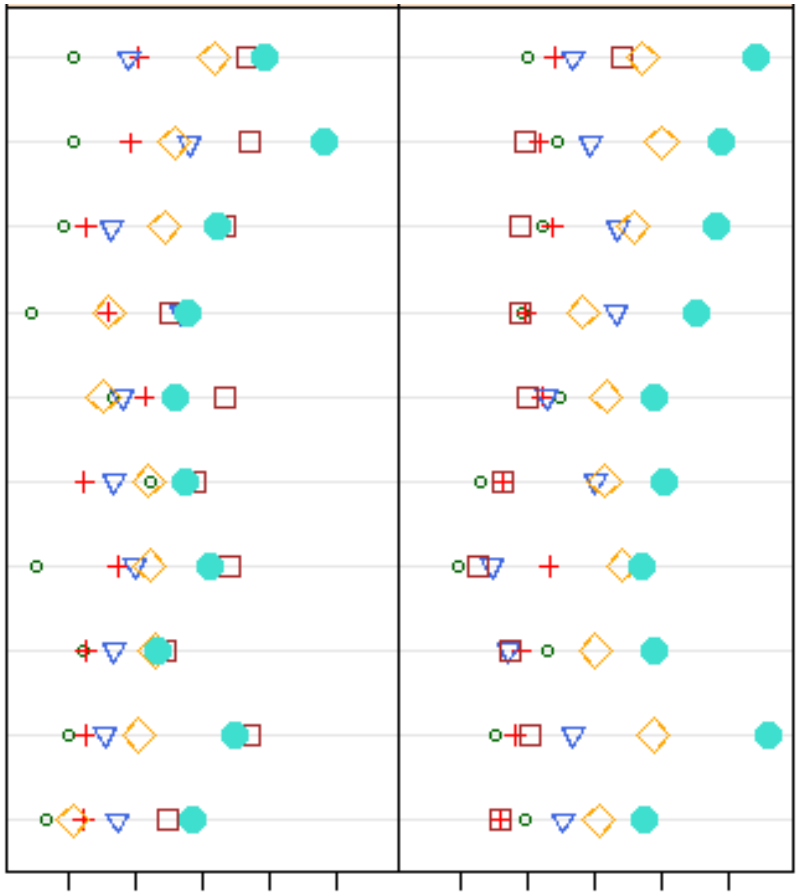
As the number of levels increase combine all these elements

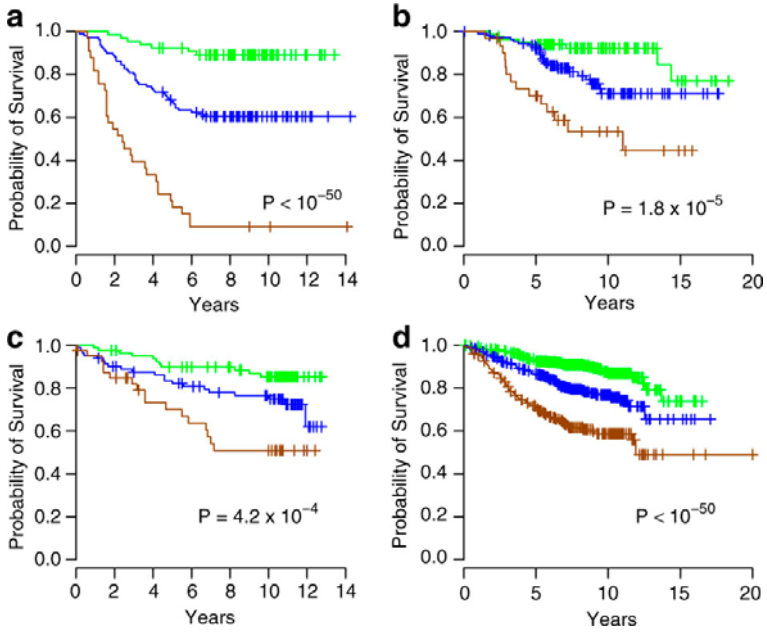
Grouped dot plots, grouped KM plots



Grouping variables

As the number of levels increase combine all these elements





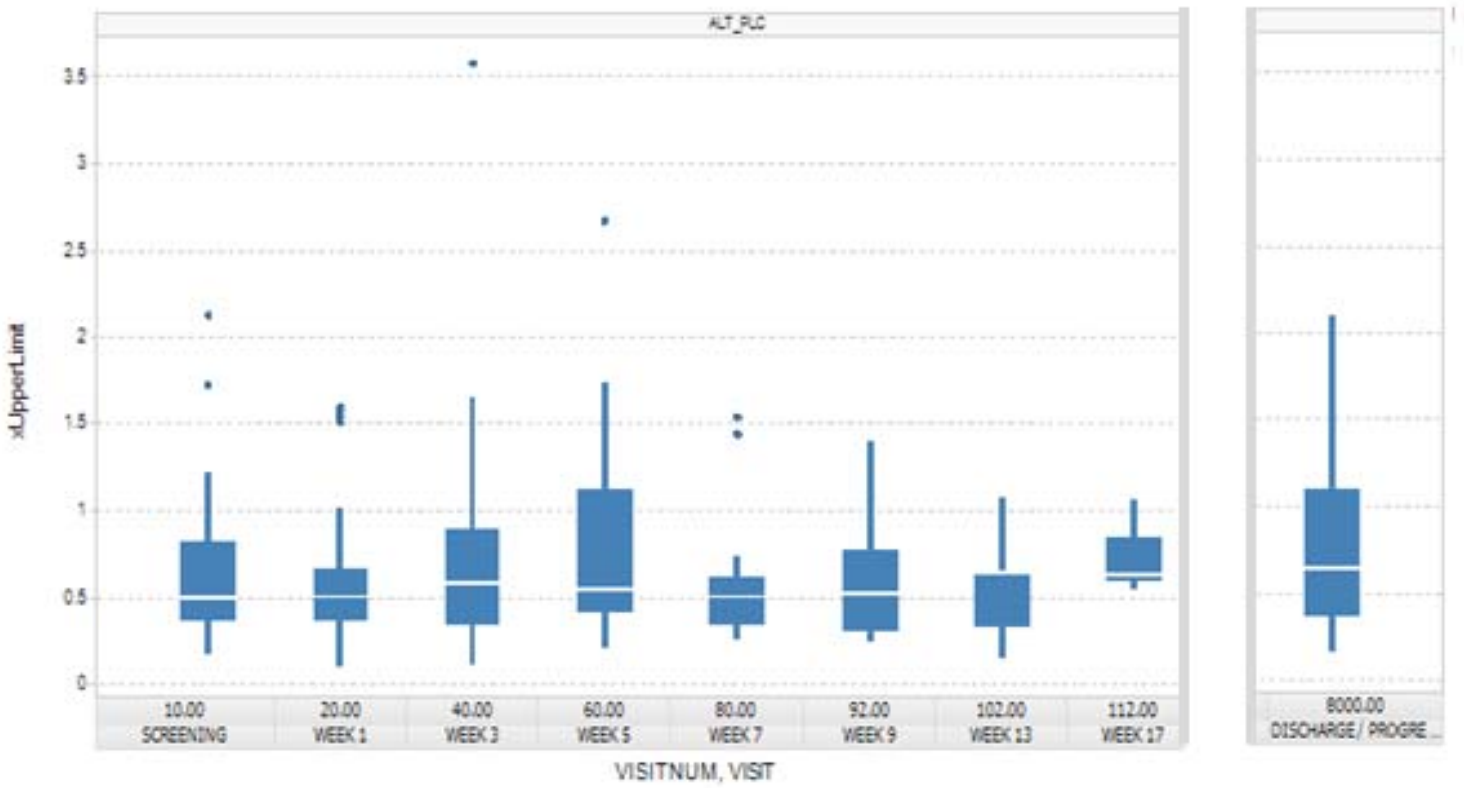
Grouping variables

Quantitative Discreet Categories

- Values have meaning
- Ordering can be utilized

Grouping variables

Quantitative Discreet Categories



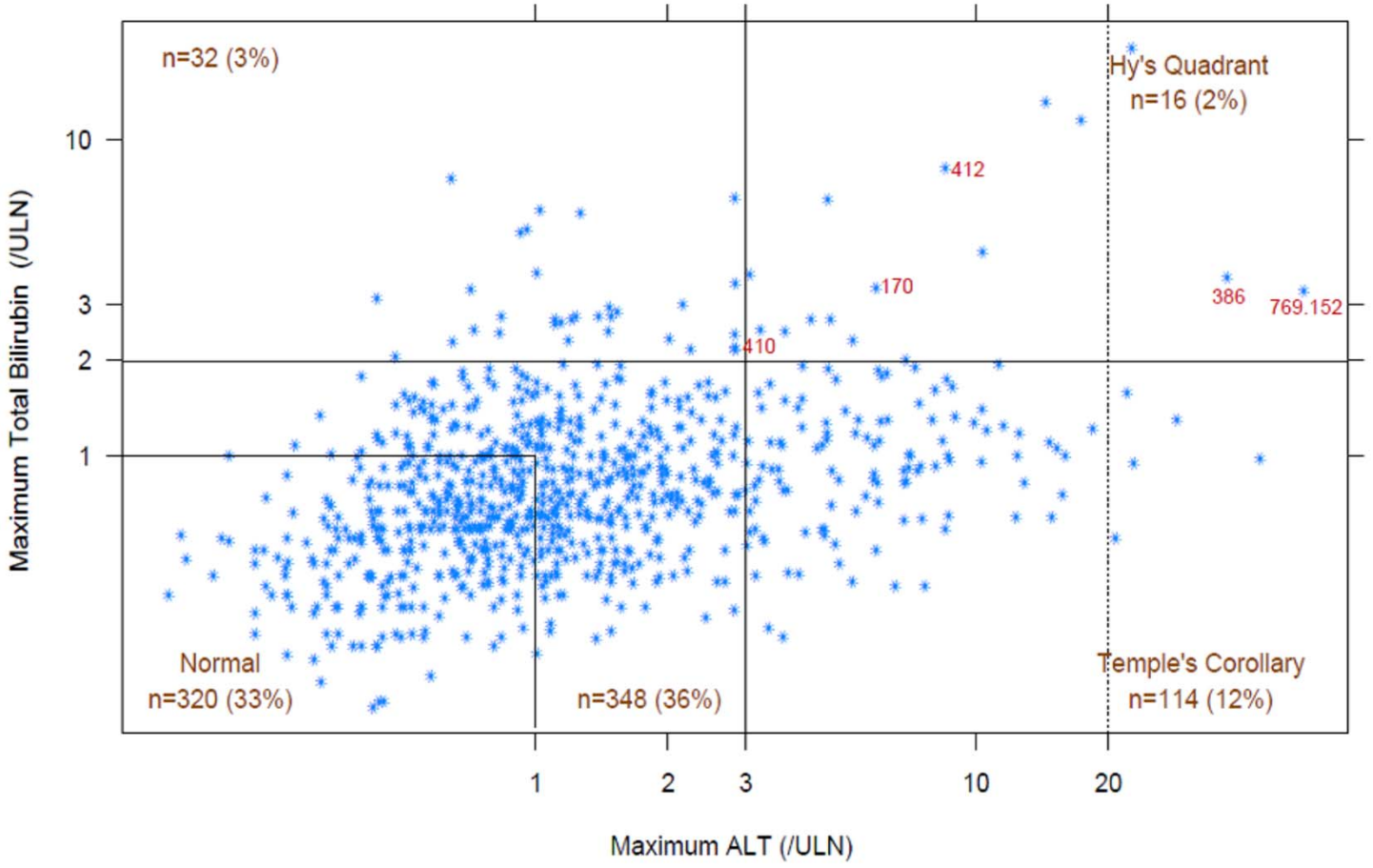
Greater than one variable of interest

Bivariate or multivariate

- E.g., maximum ALT versus maximum total bilirubin
- Examples: Scatter Plot, ellipsoid plot, hexagonal bin plots, sunflower plots, contour scatter plots

Greater than one variable of interest

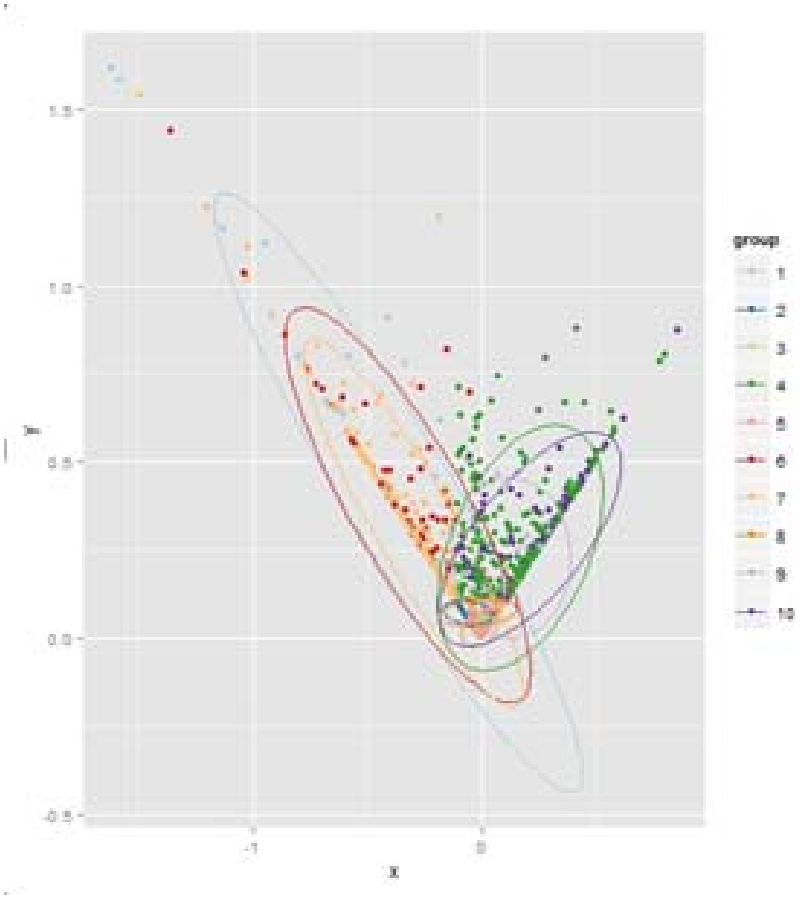
Scatter Plot – 2 Quantitative Variables of Interest



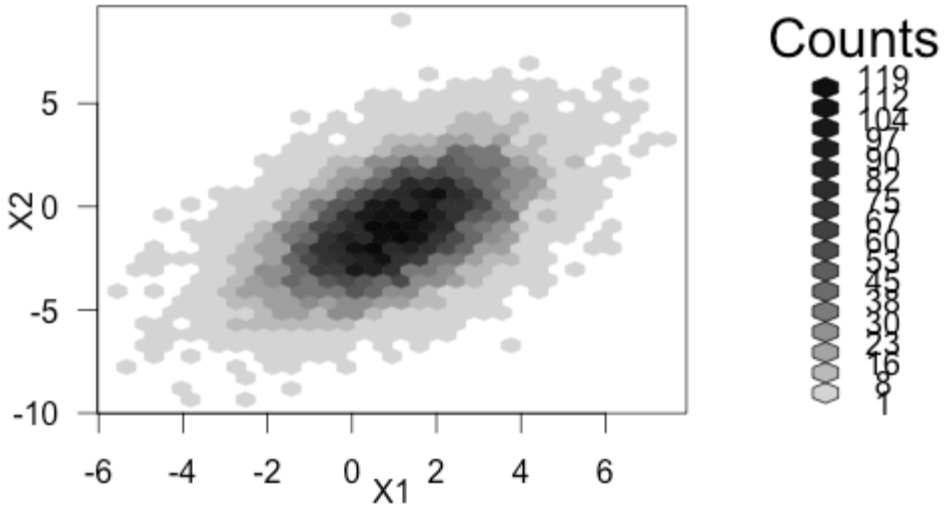
Greater than one variable of interest

Scatter Plot – 2 Quantitative Variables of Interest

– 95% CI Ellipsoid



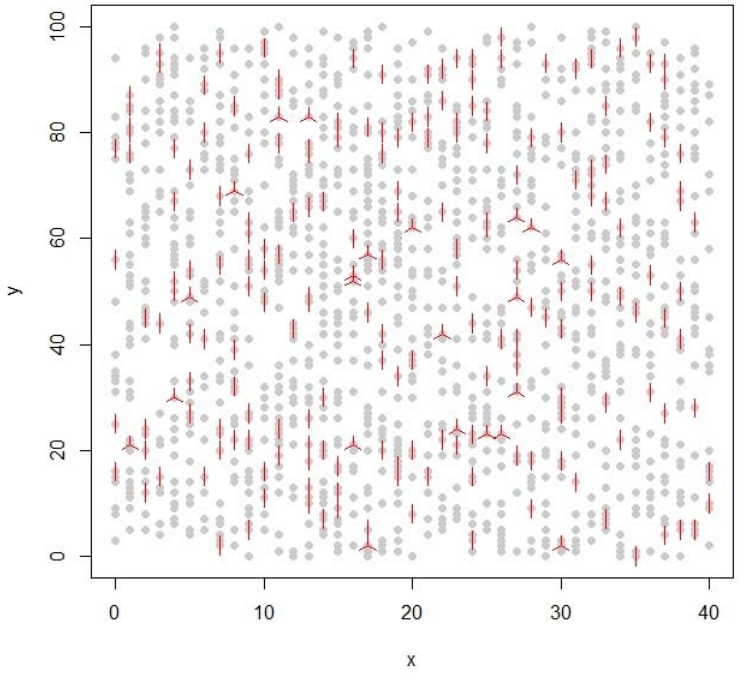
Hexagonal Bin Plots



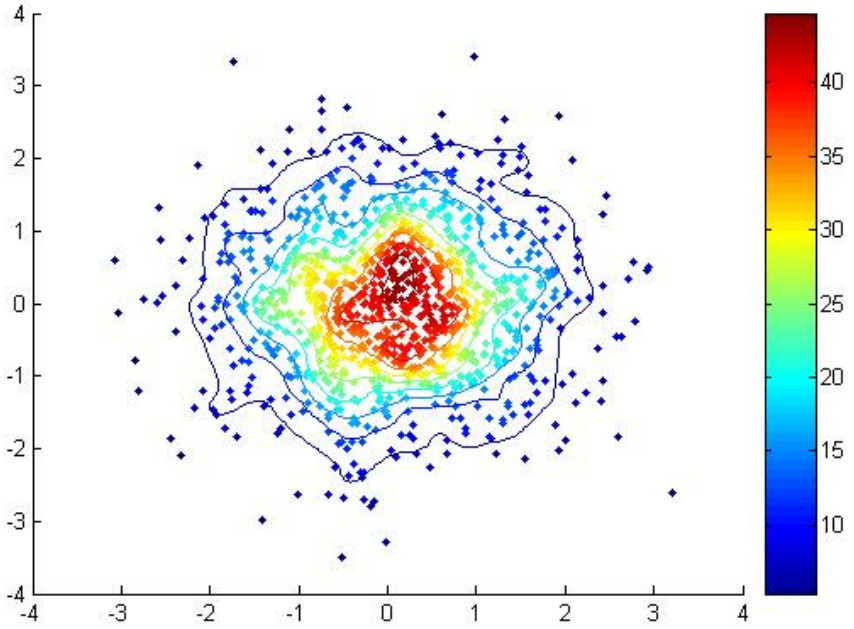
Greater than one variable of interest

Scatter Plot – 2 Quantitative Variables of Interest

– Sunflower density plot



Contour scatter matrix plot



Crowding

- As the number of elements in your figure increase graphs can become over-crowded
- Balance of bringing graphical elements and sub groups together for comparison vs. graph clutter.

Good Design Principles/Checks

Good Design Principles/Checks

- **Seeing is believing: good graphic design principles for medical research**
- *Duke SP, Bancken F, Crowe B, Soukup M, Botsis T, Forshee R. 2015. Seeing is believing: good graphic design principles for medical research. Statistics in medicine 34(22):3040–3059.*
- General Principles subteam of the Safety Graphics Working Group
- 5 good design principles or checklists
  - Tailor each graph to its primary communication purpose
  - Maximize the data-to-ink ratio
  - Bring items for comparison closer together
  - Use simplest plot for the intended purpose and audience
  - Avoid misleading visual perception

Good Design Principles/Checks

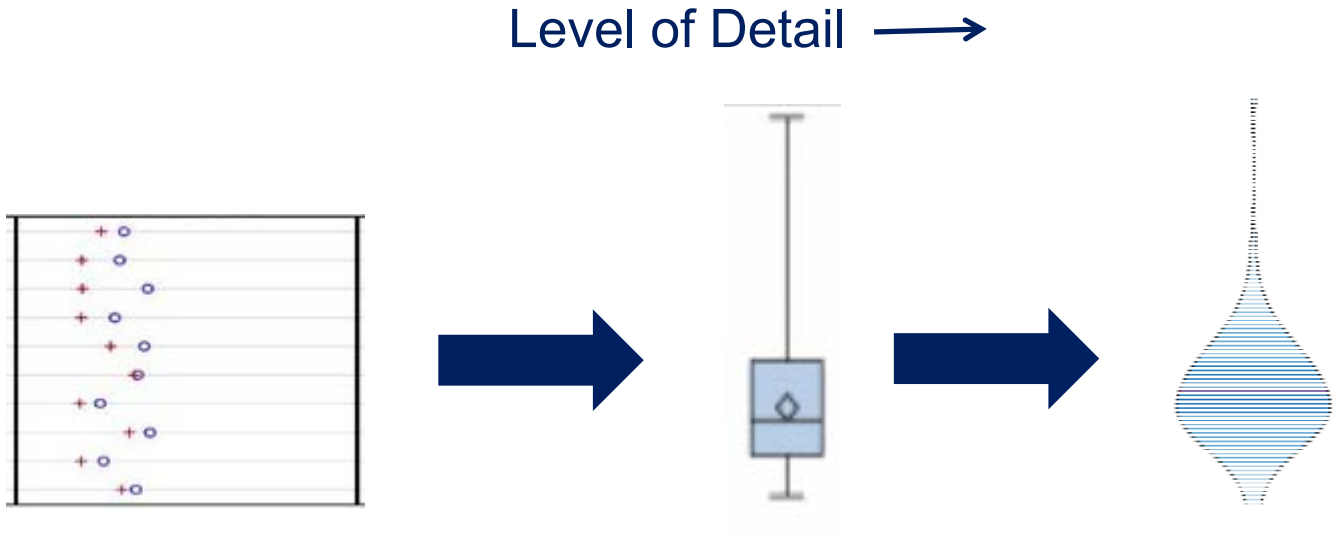
Tailor each graph to its primary communication purpose

- Each graph should stand on its own
- High level summary of results or detail statistics and distribution of the data?
  - Lower level to high level Dot plot -> box plot -> violin plot

Good Design Principles/Checks

Tailor each graph to its primary communication purpose

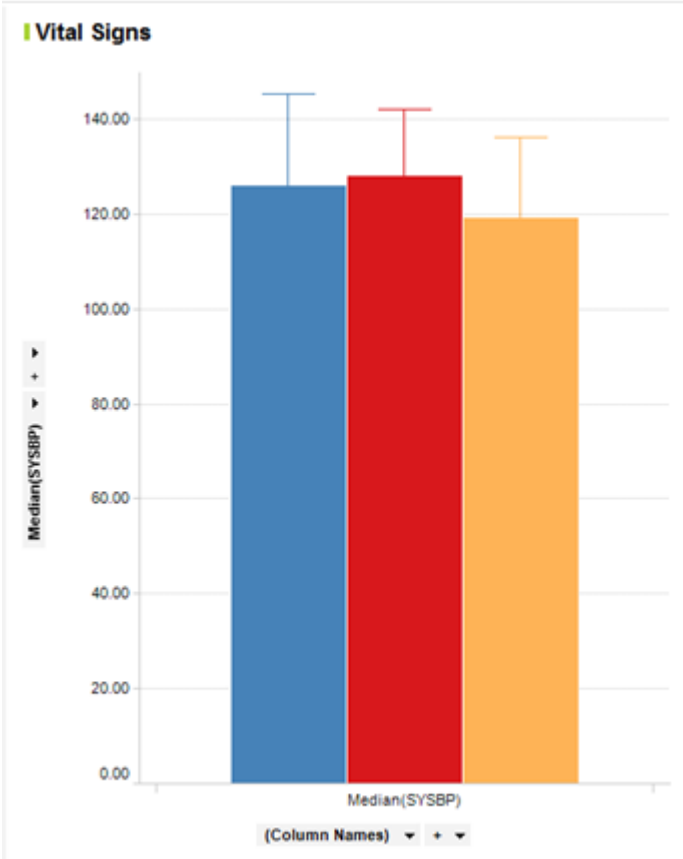
- Each graph should stand on its own
- High level summary of results or detail statistics and distribution of the data?
  - Lower level to high level Dot plot -> box plot -> violin plot



Good Design Principles/Checks

Maximize the data-to-ink ratio

- Bar plot with error bar
  - Optical illusion by creating more height.
  - Cannot see bottom of error bar – full lower to upper bound is obscured
  - Comparisons among groups are difficult without a clear picture of the uncertainty of the data
  - The height of the bar is the median (for example). The entire bar itself represents one single numerical value – data to ink ratio is not maximized



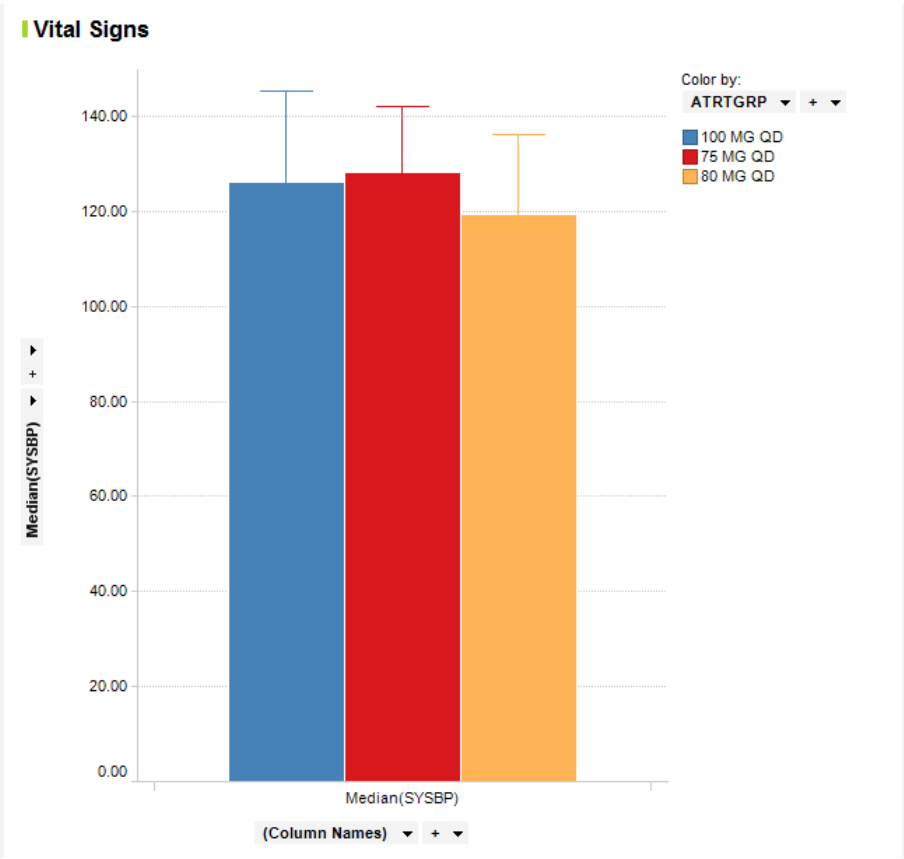
Good Design Principles/Checks

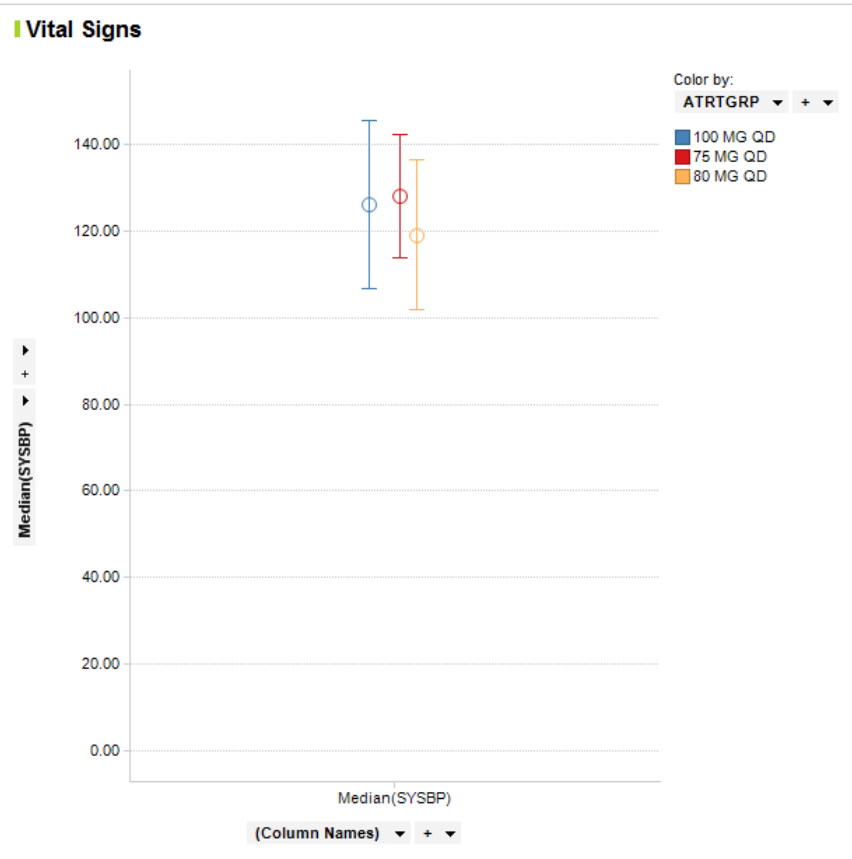
Maximize the data-to-ink ratio

- Dot plot with 95% CI
  - Allows clear picture of the uncertainty of the data
  - Single symbol depicts the median (dot) with a line representing the CI
  - Data to ink ratio is maximized
  - Increased white space results in unobstructed data patterns and higher clarity
  - Reduced ink allows for a greater number of categories to be compared

Good Design Principles/Checks

Maximize the data-to-ink ratio





Good Design Principles/Checks

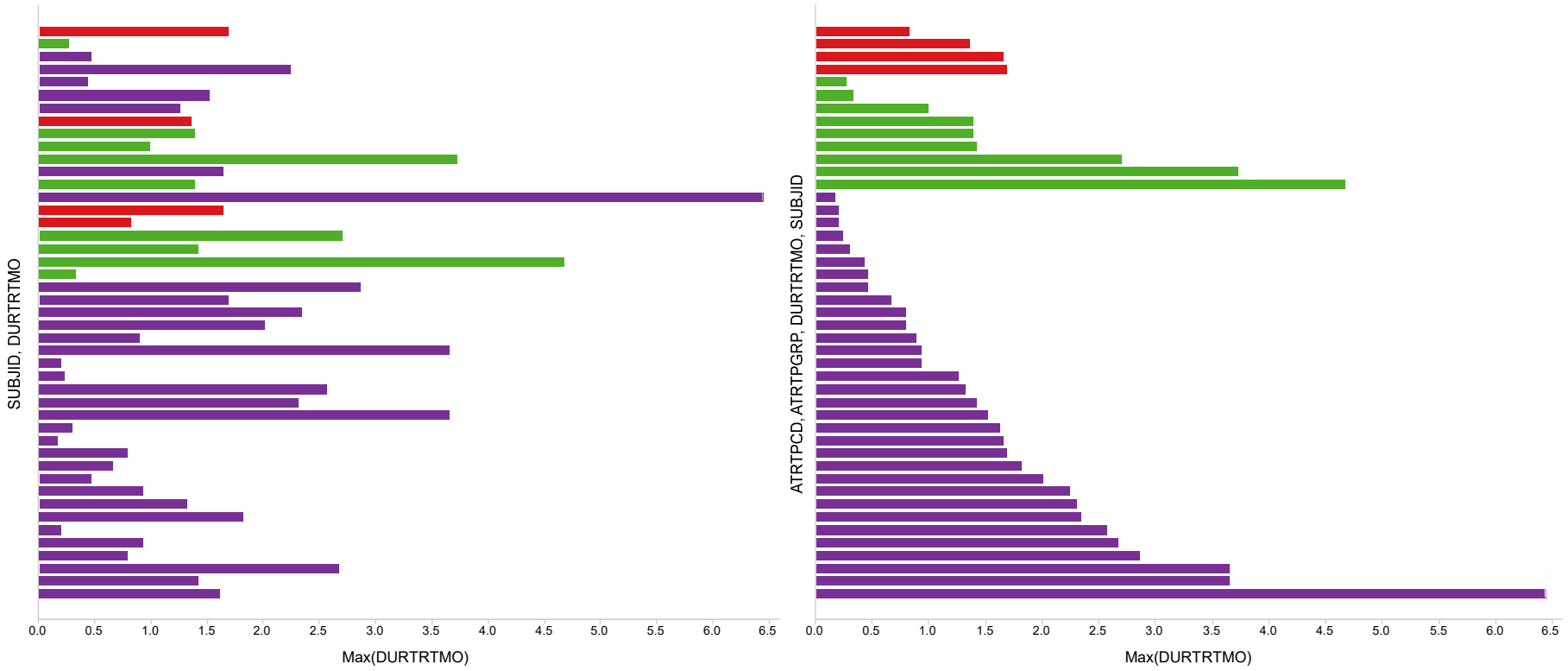
Bring items for comparison closer together

- Keep display elements as consistent as possible across displays in a set or group
- Provide and highlight common landmarks or anchors across different displays

Good Design Principles/Checks

Bring items for comparison closer together

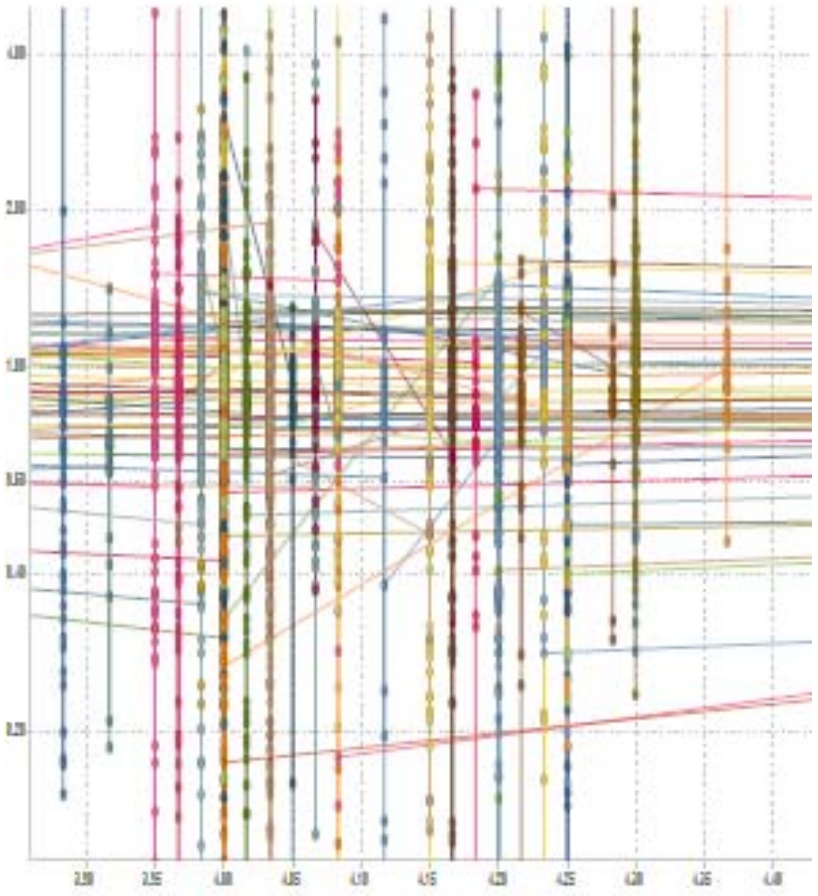
- Keep display elements as consistent as possible across displays in a set or group
- Provide and highlight common landmarks or anchors across different displays



Good Design Principles/Checks

Use simplest plot for the intended purpose and audience

- Present no more or no less information than is needed by the user
  - Relieves the user of the need to maintain detailed representation of the data in their mind
  - Too much information can lead to distraction, clutter and misinterpretation



Good Design Principles/Checks

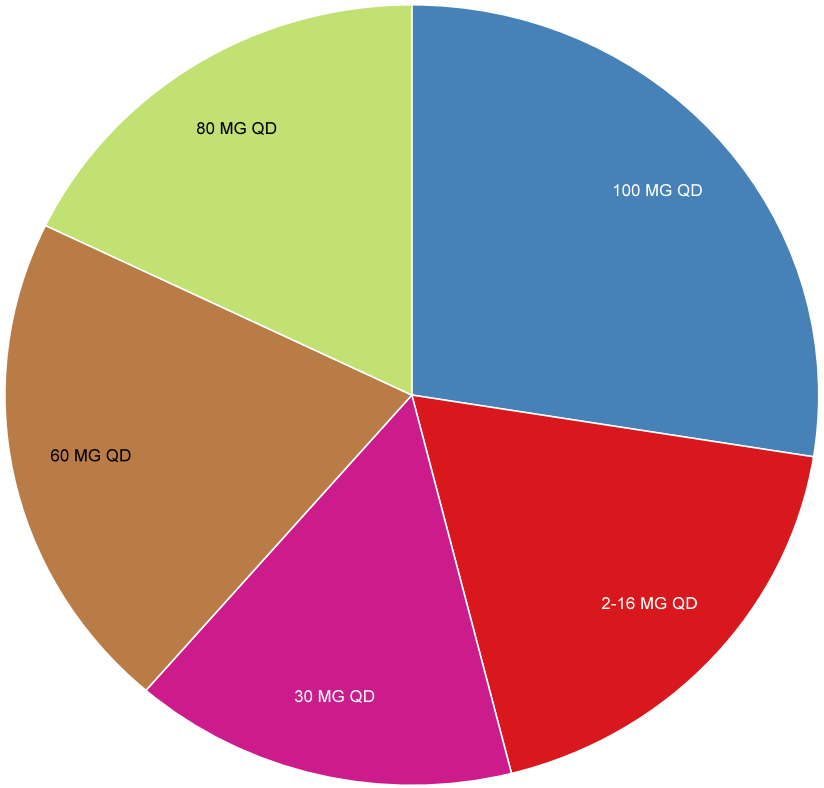
Avoid misleading visual perception

- Humans are good at identifying differences in some cases but not in others
- Some colors, angles and areas are difficult to perceive

Good Design Principles/Checks

Avoid misleading visual perception

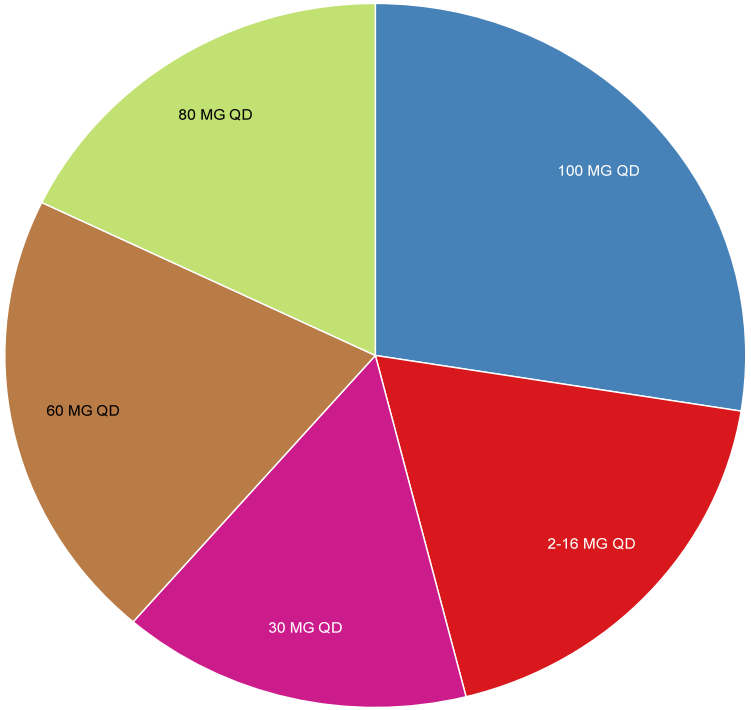
- Humans are good at identifying differences in some cases but not in others
- Some colors, angles and areas are difficult to perceive
- Pie Charts -

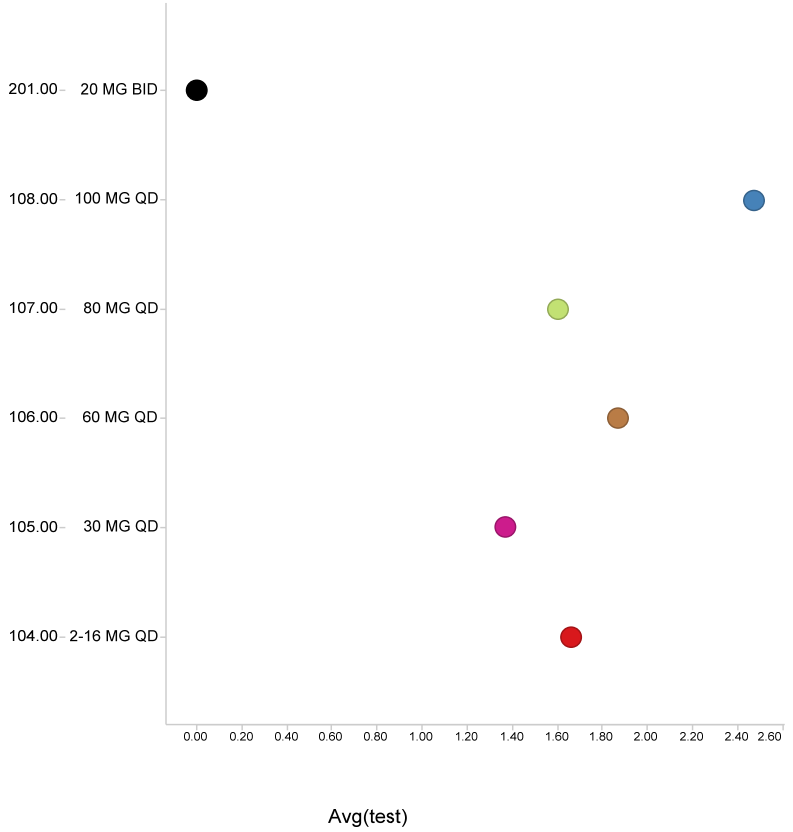


Good Design Principles/Checks

Avoid misleading visual perception

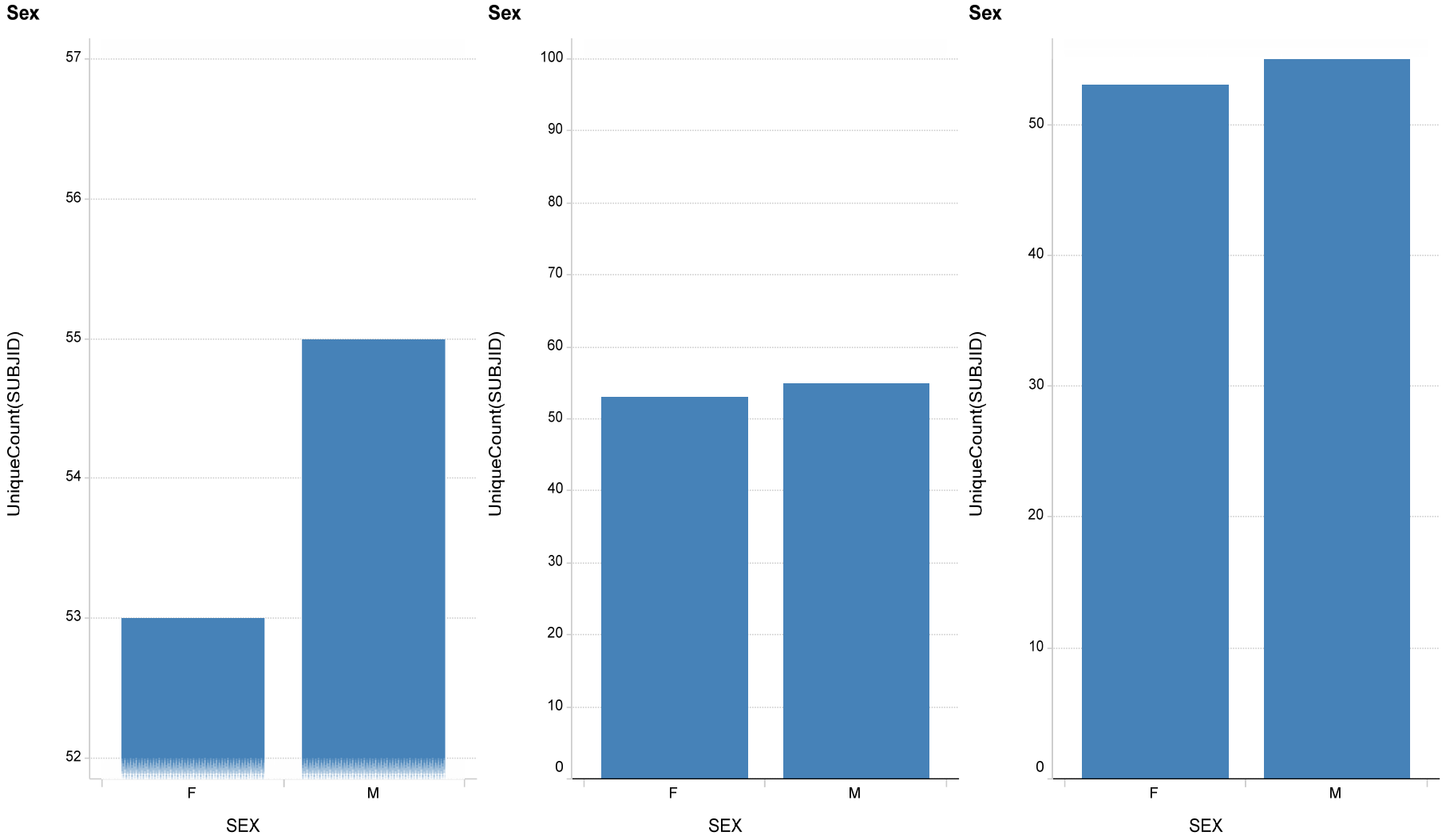
- Humans are good at identifying differences in some cases but not in others
- Some colors, angles and areas are difficult to perceive
- Pie Charts -





Good Design Principles/Checks

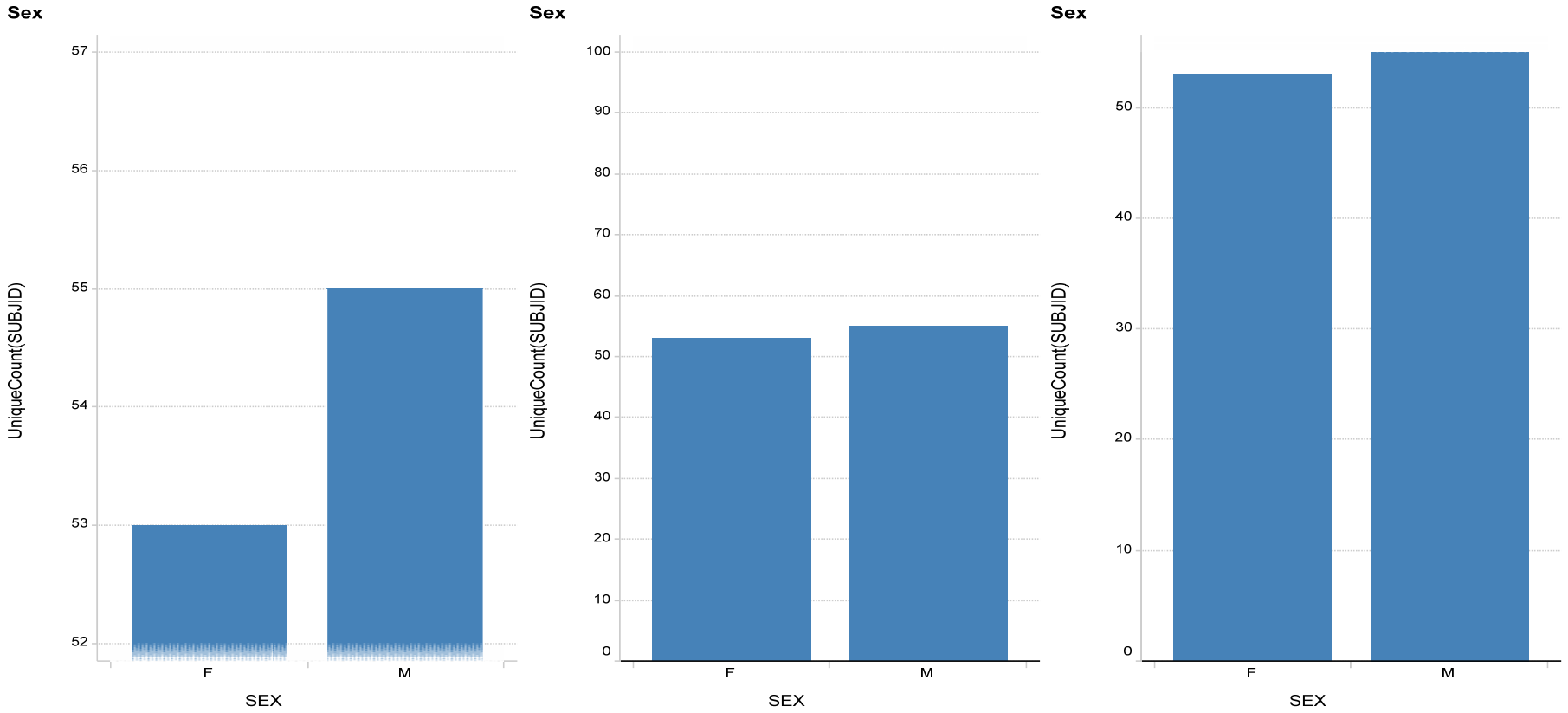
Avoid misleading visual perception – axis ranges



Good Design Principles/Checks

Avoid misleading visual perception

- Choosing axis ranges can lead to misinterpretation or misleading results
- How much change is important to capture? Make sure axes are designed for this.



Conclusion

Conclusion

- Design of effective and efficient data displays is a great skill to acquire
- The purpose of data displays is to create accurate and effective communication for decision makers
- Data can be displayed in text, table or graphical format. Each type has pros and cons and will be more appropriate in certain circumstances
- When choosing what type of figure, the choice of graph will depend on the data type of the variable(s) of interest
- For quantitative variables, the choice of graph depends on the desired amount of detail about the distribution of the underlying data
- For character variables, the choice of graph depends on the number of levels within the category
- The number of subgroups for comparison will also dictate graph choice
- Develop of set of good graphical principles to help check the quality of your output
- Peer-review and revise the displays for the highest quality

Thank You