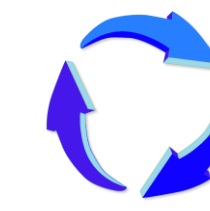


# Show the data instead of a bar graph for continuous data

Author Tracey Weissgerber (0000-0002-7490-2600), PhD, Mayo Clinic, QUEST, BIH Center for Transforming Biomedical Research

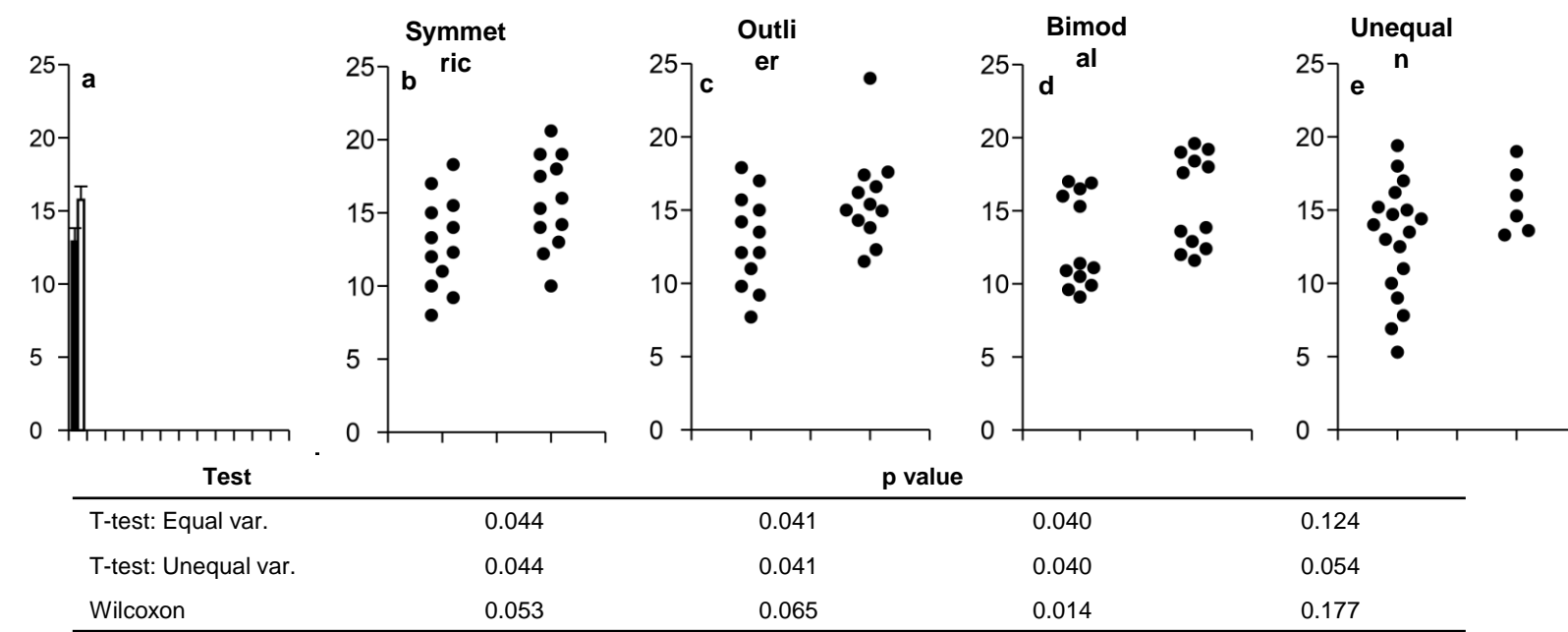


Replicate  
Reproduce  
Reuse

eLIFE Ambassadors 2019 CC-BY 4.0

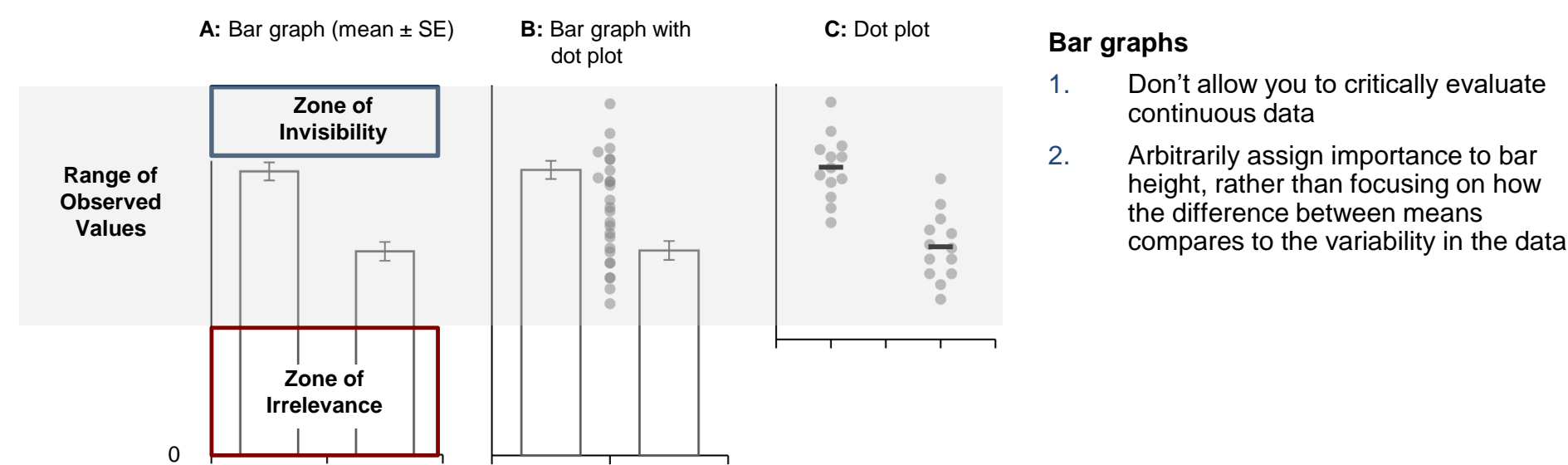
International license.

Why shouldn't I use a bar graph for continuous data?  
Different datasets can lead to the same bar graph:



The data may suggest different conclusions from the summary statistics.

Can I use a bar graph even if my data are normally distributed?  
Not really, because:



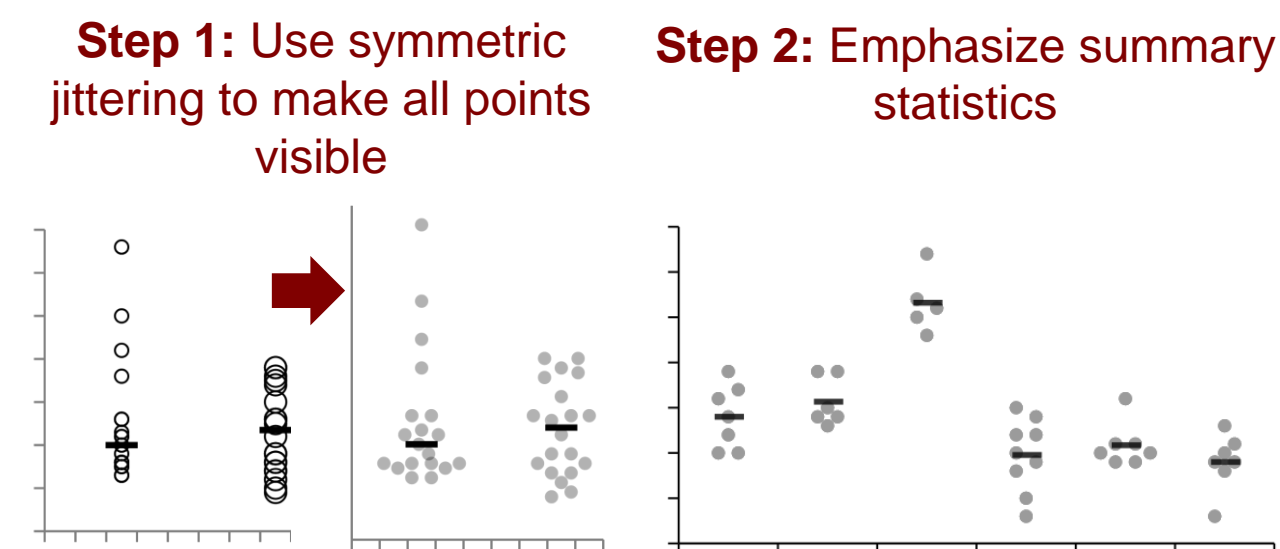
- Bar graphs**
- Don't allow you to critically evaluate continuous data
  - Arbitrarily assign importance to bar height, rather than focusing on how the difference between means compares to the variability in the data

Do I need expensive software?  
No, there are free, user-friendly tools:

- Gnuplot
- Veusz
- Augl
- Candela
- Chartblocks
- R

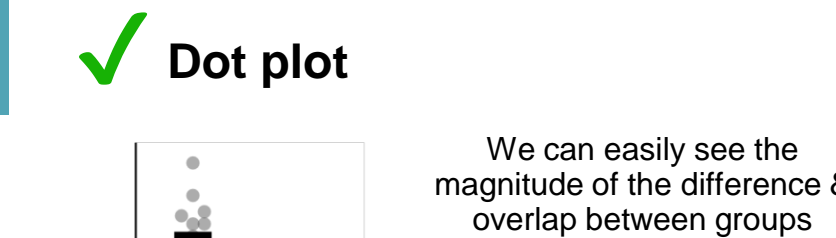
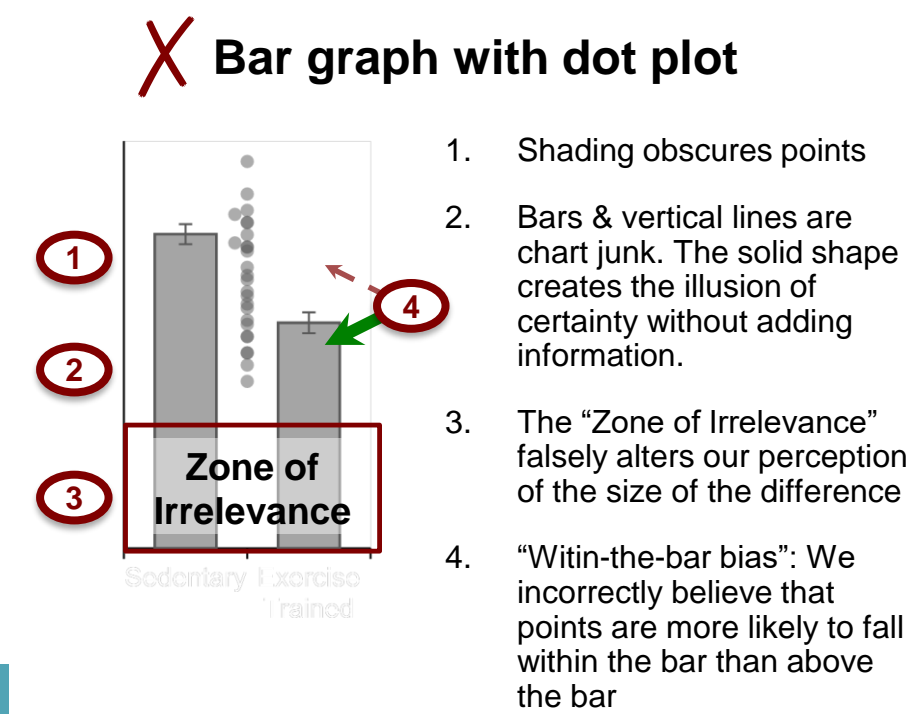
...but bar graphs convey a clear message. Sometimes it's hard to see what's going on with dot plots.

Emphasize summary statistics & de-emphasize data points to convey a clear message while allowing readers to critically evaluate the data



Can I just add data points to my bar graph?

Why dot plots are better than bar graphs with points



Why shouldn't I use box plots for bimodal data?

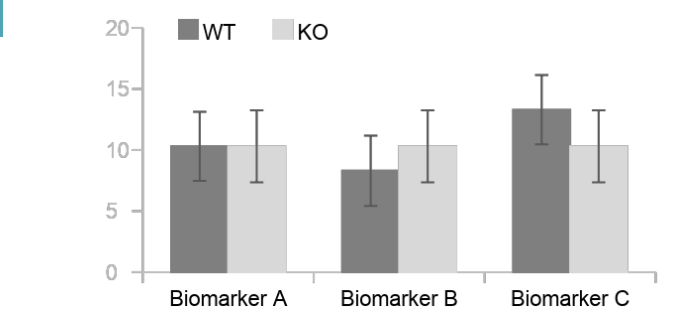
Box plots conceal the two peaks. Bimodal distributions are easier to see with small (dot plots) or larger (violin plots) samples

Does it matter how I did my analysis?

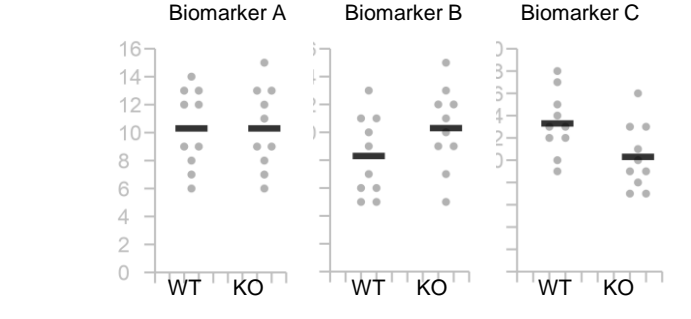
Yes! The figure structure gives the reader information about your design & analysis. Avoid sending mixed messages, especially with small datasets.

**Experimental goal:** Compare wild type vs. knockout mice  
**Statistical analysis:** t-tests were used to compare values for each dependent variable (biomarker A, B and C).

**X Sending Mixed Messages**  
Figure structure erroneously suggests that authors also intended to compare biomarkers A, B and C



**Clear communication**  
Figure structure matches study design & analysis, shows that the authors did not intend to compare biomarkers

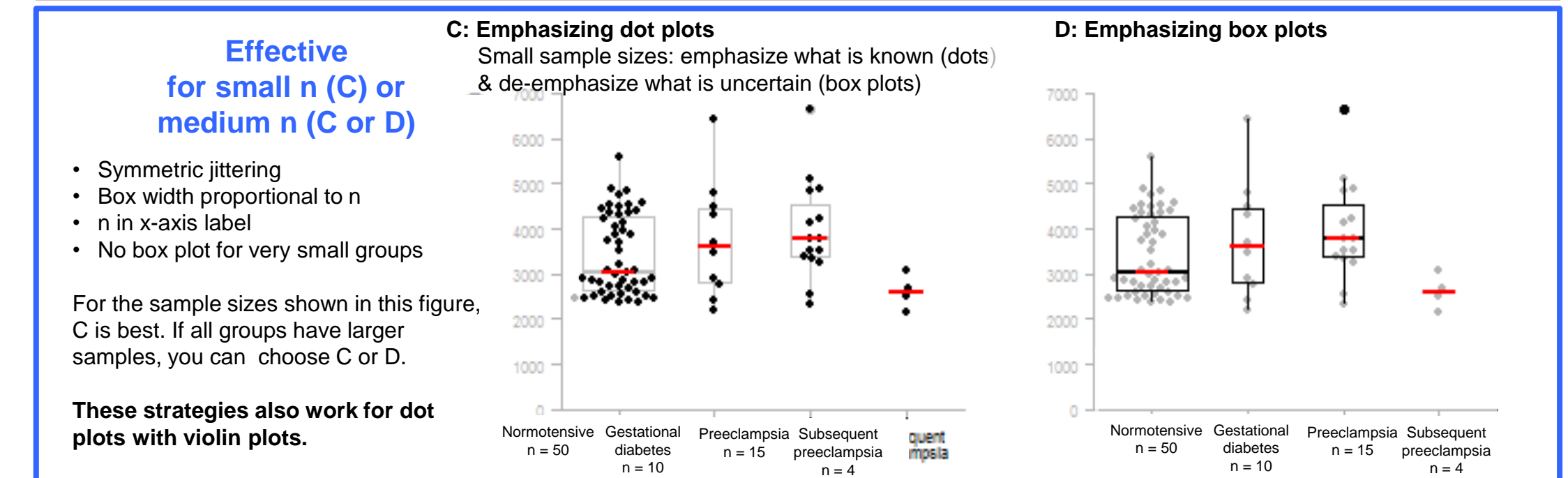
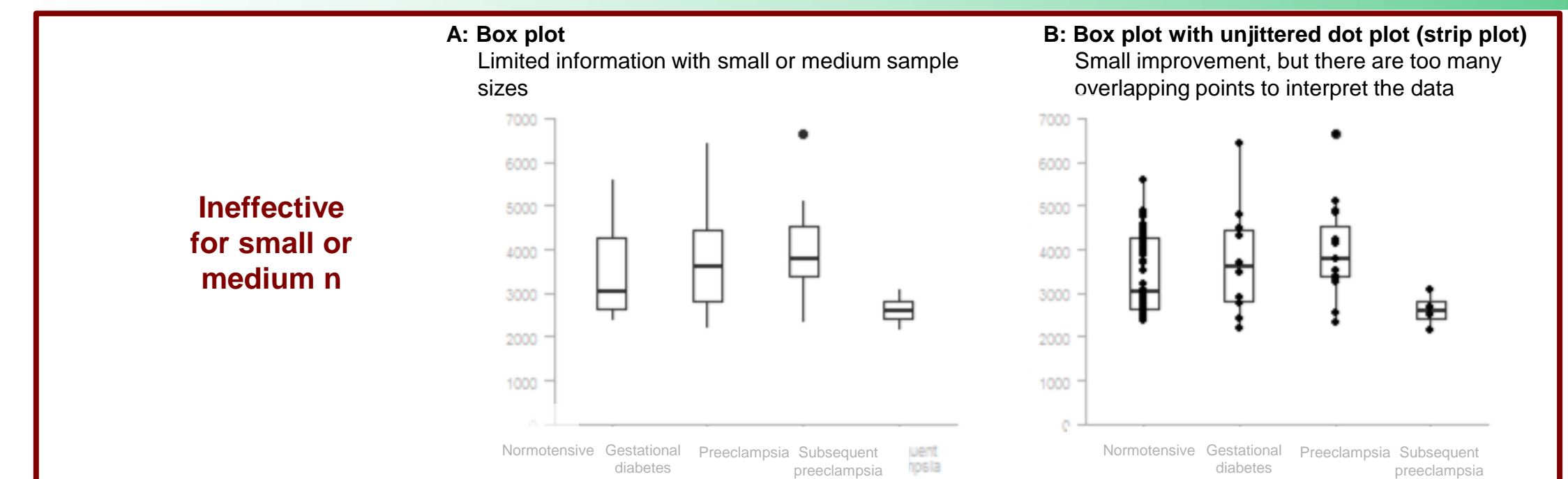


How do I design figures that match my study design & analysis?

For simple analyses of small datasets, it's often clearest to show one graph per analysis that includes all groups, time points or conditions in the analysis.

Analysis Strategy	Example	Figure Structure	Illustration
Comparing groups	Figure compares wild type vs. knockout mice	One figure showing all groups that were included in the analysis	
Repeating the same analysis on different dependent (outcome) variables	Figure compares wild type vs. knockout mice. Three different tests are performed on different biomarkers.	Separate panels for each analysis (i.e. dependent variable)	
Comparing groups with pooled subgroups	Figure compares wild type vs. knockout mice. Male and female mice are pooled.	One figure showing all groups that were included in the analysis; data points for different subgroups are shown in different colors	
Stratified analysis	Figure compares wild type vs. knockout mice. Separate analyses are performed for males and females.	Separate panels for each analysis When possible, using the same scales can facilitate visual comparisons	
Testing for an interaction	Figure compares four different groups of mice (wild type / female, knockout / female, wild type / male, knockout / male). The analysis tests for an interaction between strain and sex.	One figure showing all groups included in the analysis	

How do I make effective box or violin plots with dot plots?

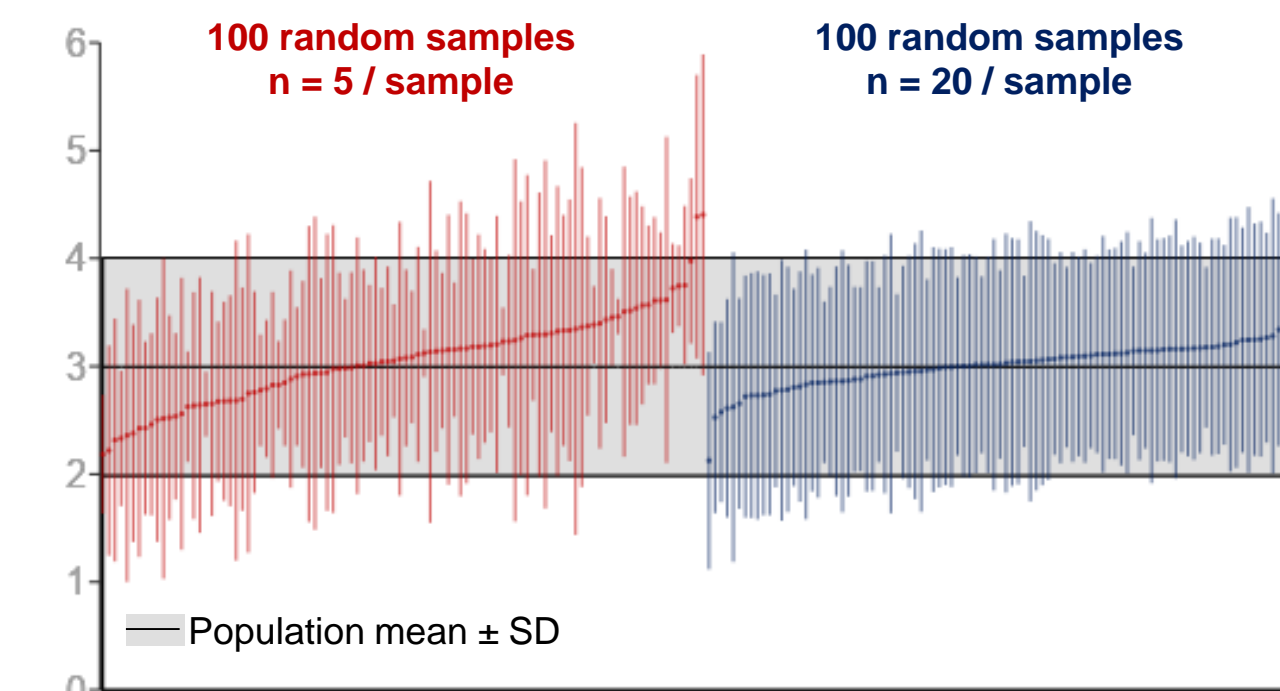


Plots to use for normally distributed, continuous data:

Figure Types	Example	Type of Outcome Variable	What the Plot Shows	Sample Size	Data Distribution	Best Practices
Dot plot		Continuous	Individual data points & mean or median line Other summary statistics (i.e. error bars) can be added for larger samples	Very small OR small; can also be useful with medium samples	Sample size is too small to determine data distribution OR Any data distribution	Make all data points visible - use symmetric jittering Many groups: Increase white space between groups, emphasize summary statistics & de-emphasize points Only add error bars if the sample size is large enough to avoid creating a false sense of certainty
Dot plot with box plot or violin plot		Continuous	Combination of dot plot & box plot or violin plot (see descriptions above and below)	Medium	Any	Make all data points visible (symmetric jittering) Smaller n: Emphasize data points and de-emphasize box plot, delete box plot and show only median line for groups with very small n Larger n: Emphasize box plot and de-emphasize points
Box plot		Continuous	Horizontal lines on box: 75 <sup>th</sup> , 50 <sup>th</sup> (median) and 25 <sup>th</sup> percentile Whiskers: varies; often most extreme data points that are not outliers Dots above or below whiskers: outliers	Large	Do not use for bimodal data	List sample size below group name on x-axis Specify what whiskers represent in legend
Violin plot		Continuous	Gives an estimated outline of the data distribution. The precision of the outline increases with increasing sample size.	Large	Any	List sample size below group name on x-axis
Bar graph		Counts or proportions	Bar height shows the value of the count or proportion	N/A	N/A	Do not use for continuous data

When is it appropriate to show summary statistics like error bars, box plots or violin plots?

Summary statistics are only meaningful when there are enough data to summarize...



With small samples, means and SDs can be very different from the true population values...