Comet analysis – Issues, Limitations and Recommendations

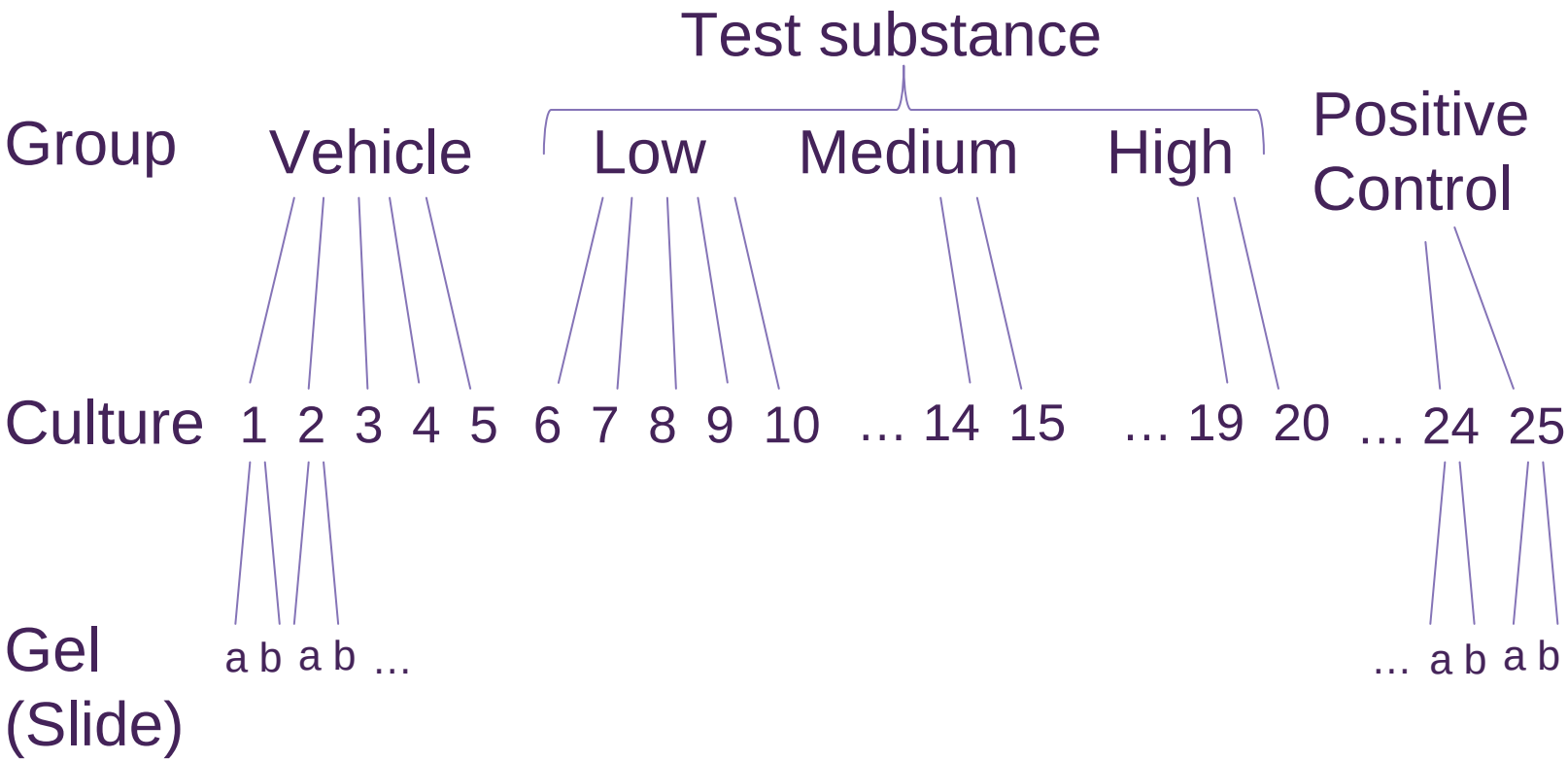
Jonathan Bright

Discovery Statistics

AstraZeneca

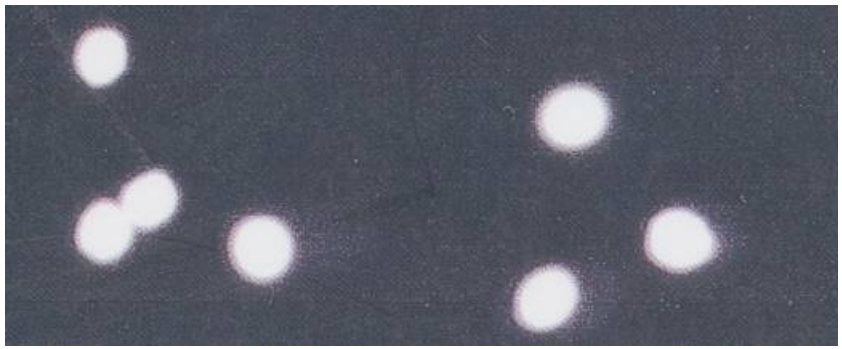
AstraZeneca
life inspiring ideas

Nested Design

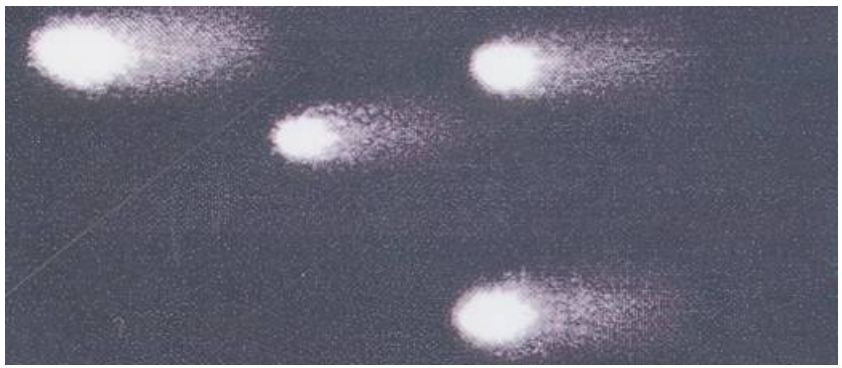


Measurements on 50 cells per gel

Undamaged and Damaged Cells



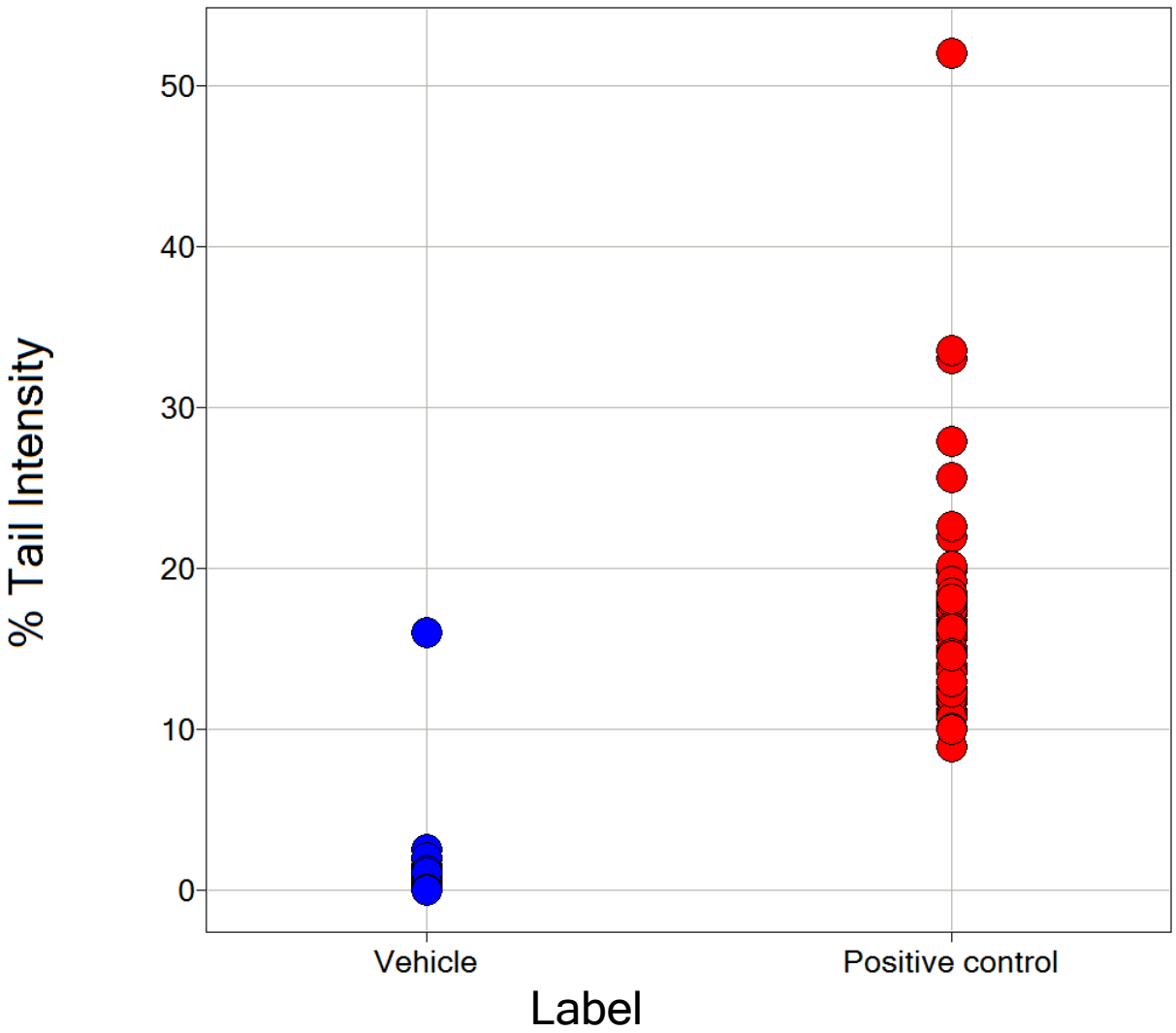
Vehicle-treated cells showing no genetic damage



Positive control-treated cells showing severe genetic damage

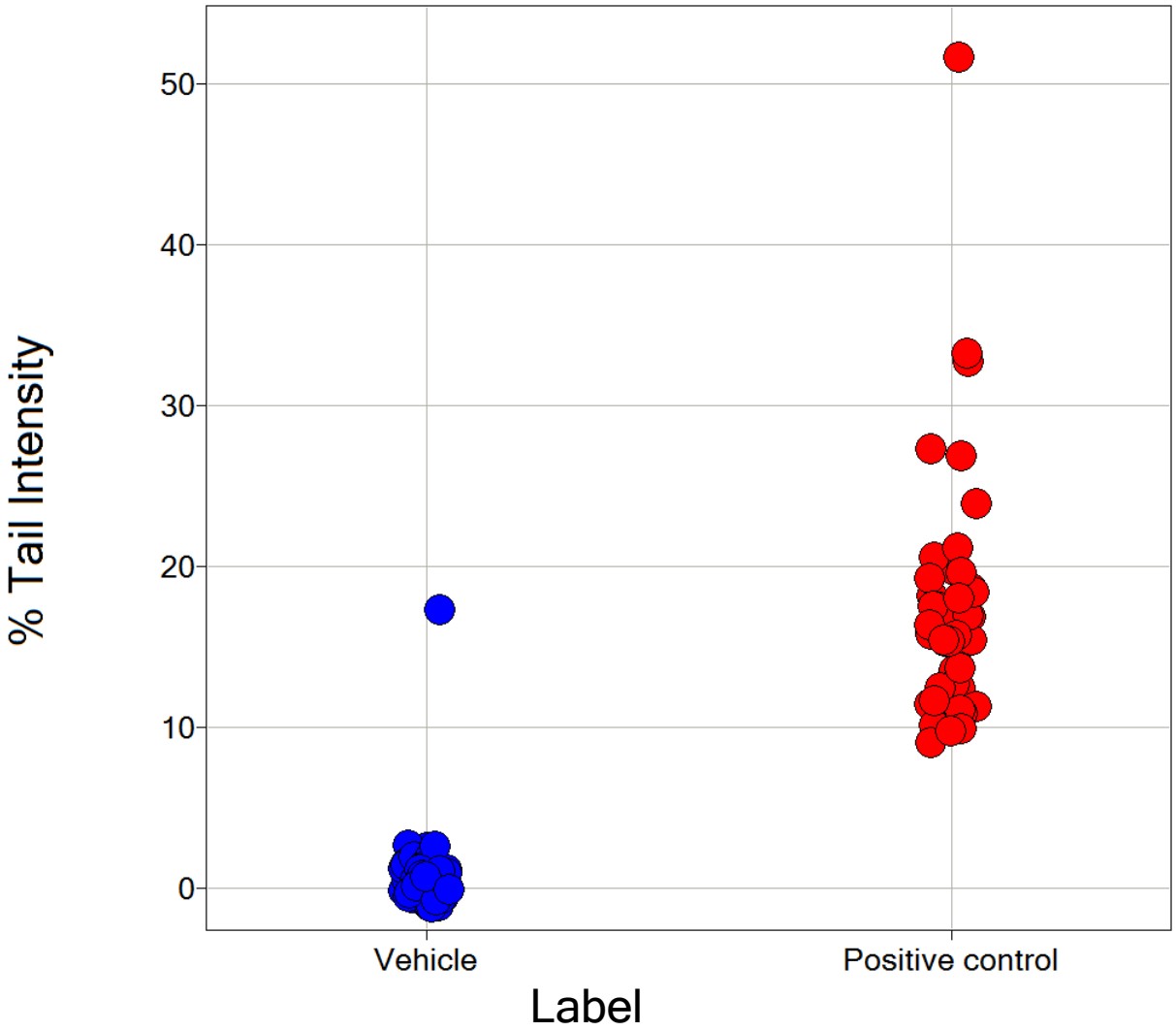
Jonathan Bright, Comet Analysis – Issues, Limitations and Recommendations, NCSC 2010

Tail Intensities (linear scale)



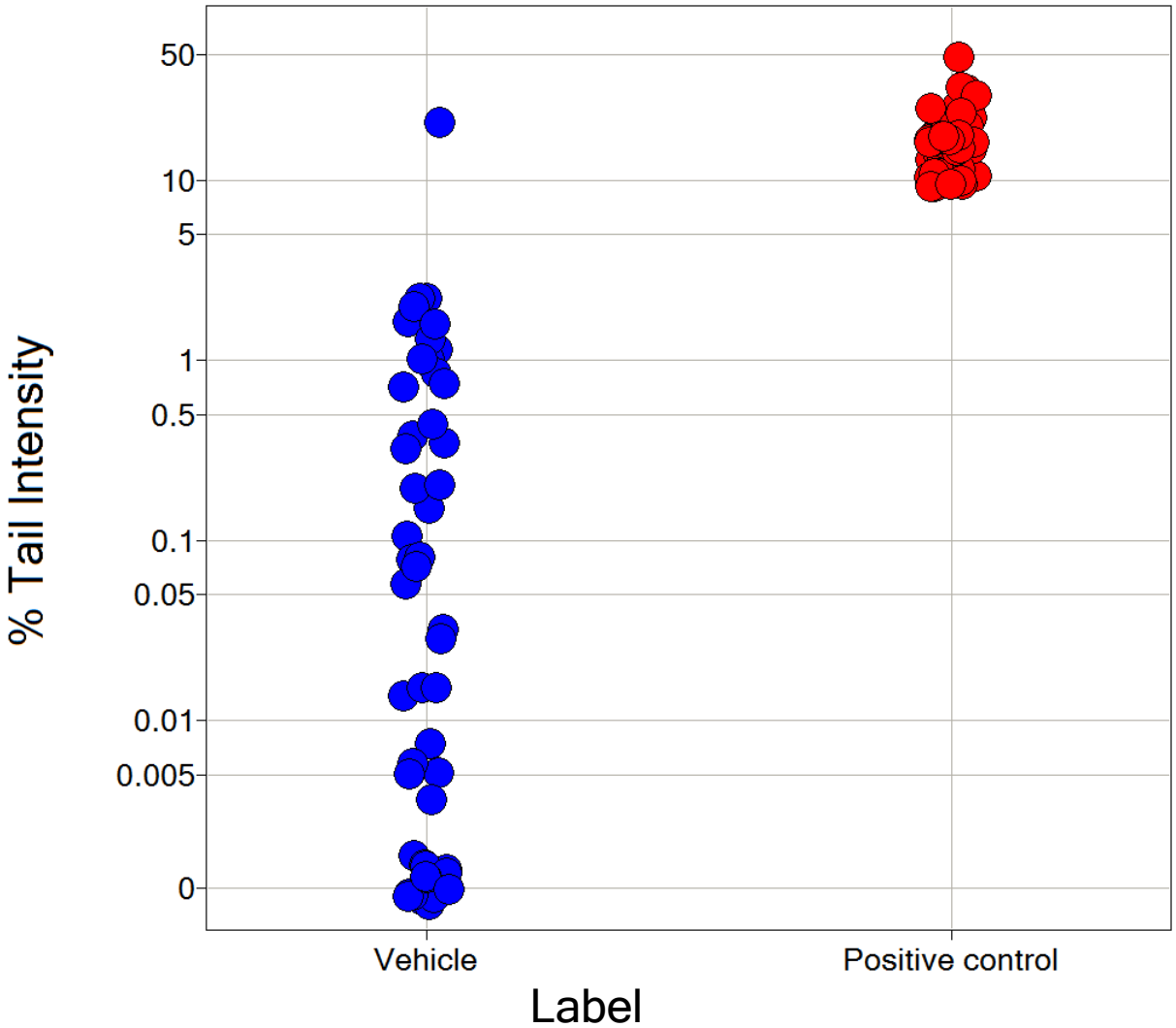
Jonathan Bright, Comet Analysis – Issues, Limitations and Recommendations, NCSC 2010

Tail Intensities (linear, jittered)



Jonathan Bright, Comet Analysis – Issues, Limitations and Recommendations, NCSC 2010

Tail Intensities (log scale, jittered)



Jonathan Bright, Comet Analysis – Issues, Limitations and Recommendations, NCSC 2010

Summary and Analysis

- Summarise each gel of 50 numbers with a statistic, S
- Analyse S
  - Excluding positive control data
  - Using PROC MIXED
  - Pairwise contrasts
- Main issue is the choice of gel summary, in particular when:
  - Large number of zeros
  - Small number of unusually large values

Summary Statistic 1. Mean Log

- Non-zero part of distribution approx lognormal
- Mean is powerful
- Fold-change treatment effects

BUT

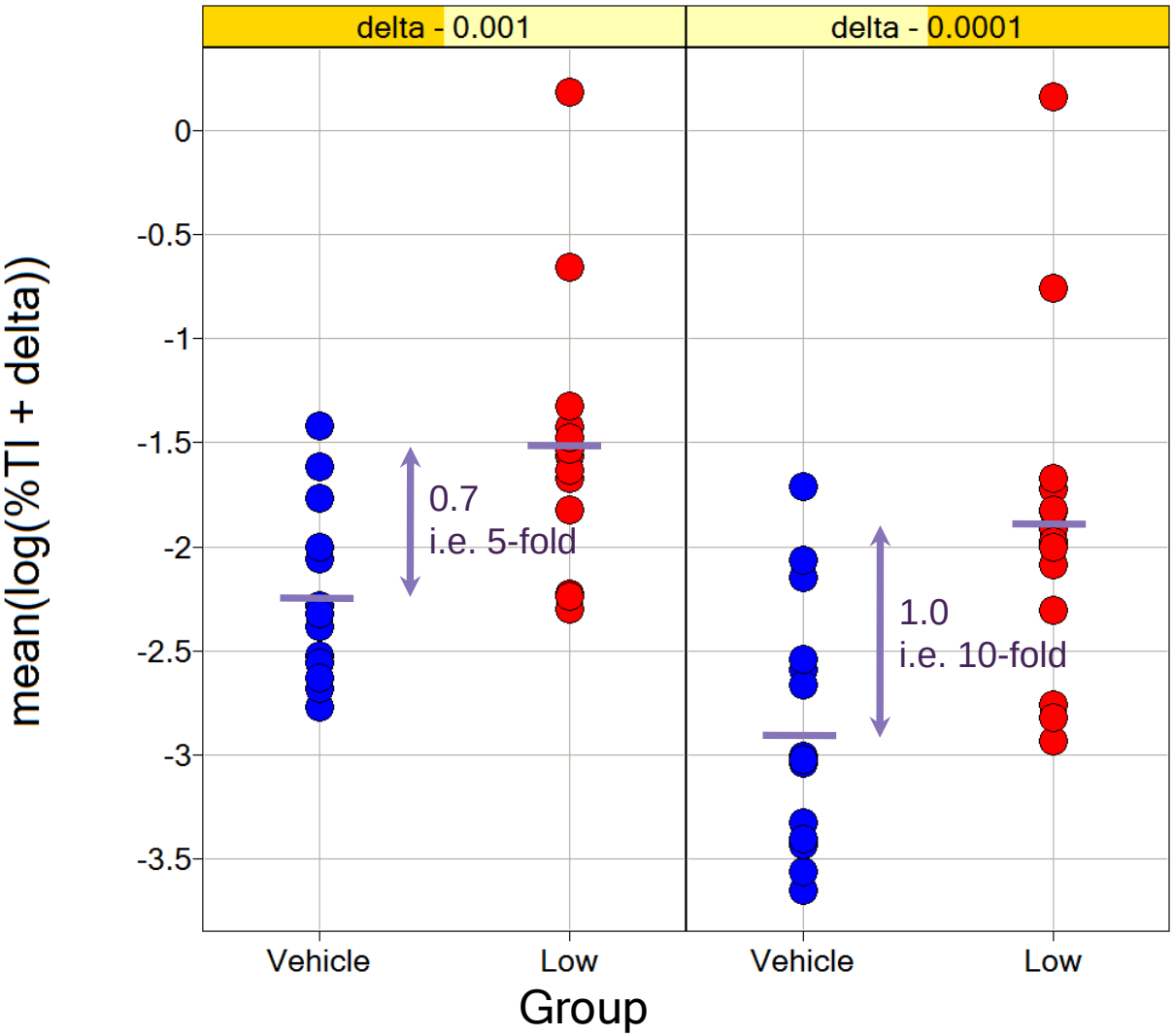
- Zeros!
- Add delta (e.g. 0.001) to all tail intensities before logging and averaging

Summary Statistic 1. Mean Log (contd)

- 50 zeros on a gel
  - S = mean(log(tail intensity + delta)) = log(delta)
  - Delta = 0.001 then S = -3
  - Delta = 0.0001 then S = -4 etc
  - i.e. S depends critically on delta
- 50 tail intensities all > 0.1, say
  - S approx = mean(log(tail intensity))
  - i.e. S is approx independent of delta
- Treatment effect depends on delta!

AstraZeneca Discovery Statistics

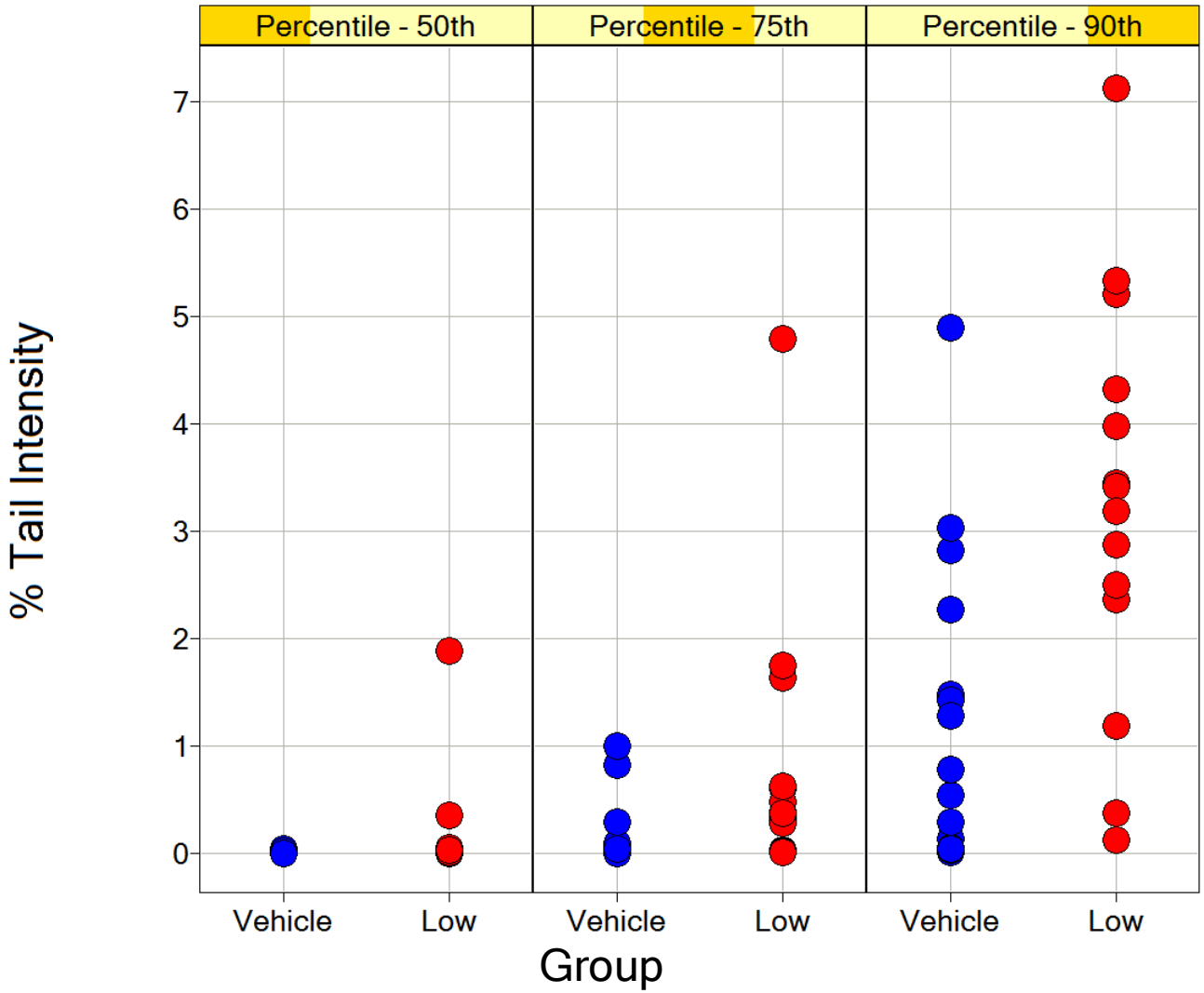
Summary Statistic 1. Example



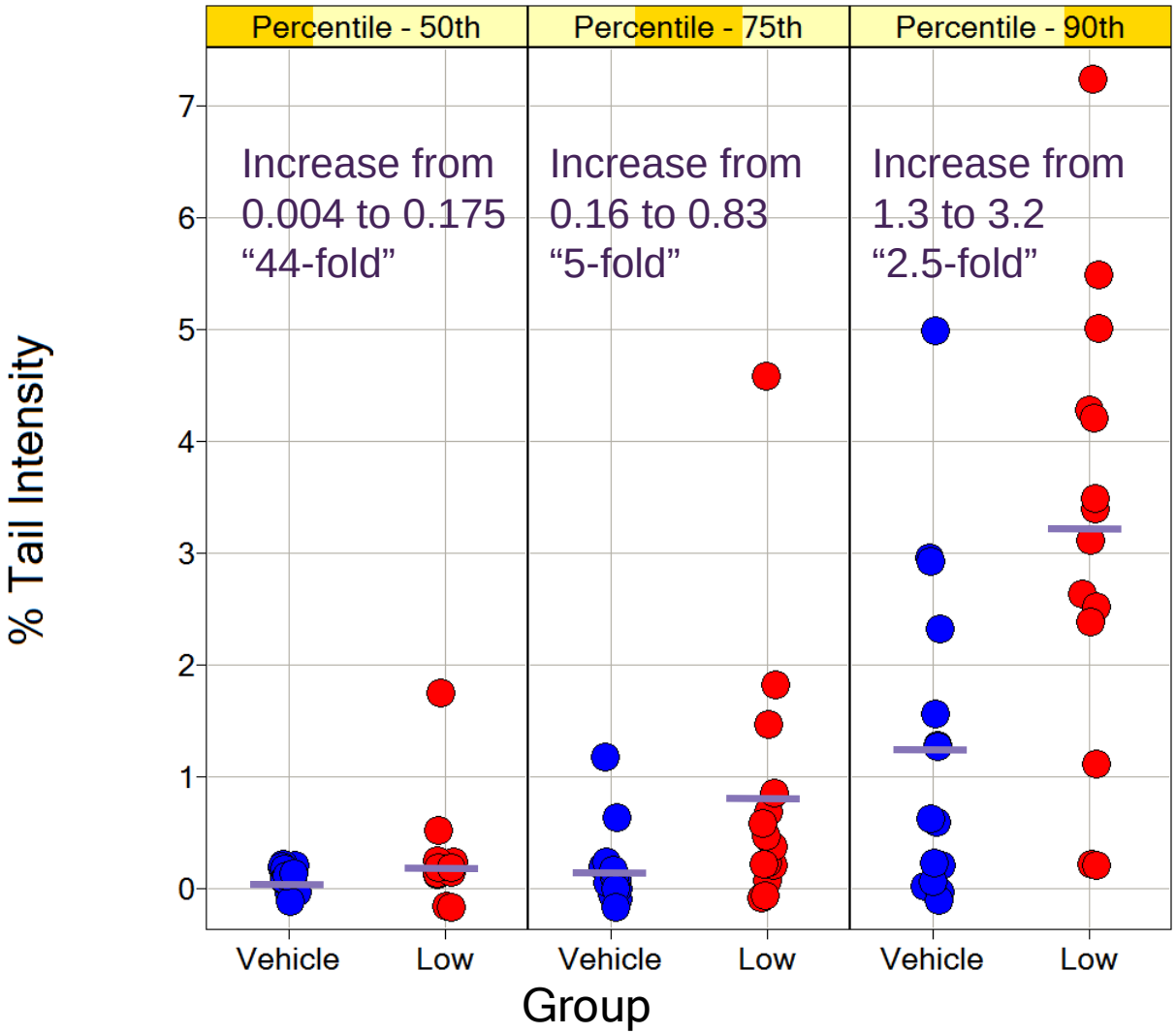
Summary Statistic 2. Percentiles

- Median
  - Robust measure of location
  - Will fail to detect changes of interest in the upper tail
- 90th percentile
  - Better chance of detecting changes in the upper tail
  - Not very robust with only 50 values
- 75th percentile
  - May offer a better balance than either of the above
- (In the presence of many zeros, even the chosen percentile may equal zero.)

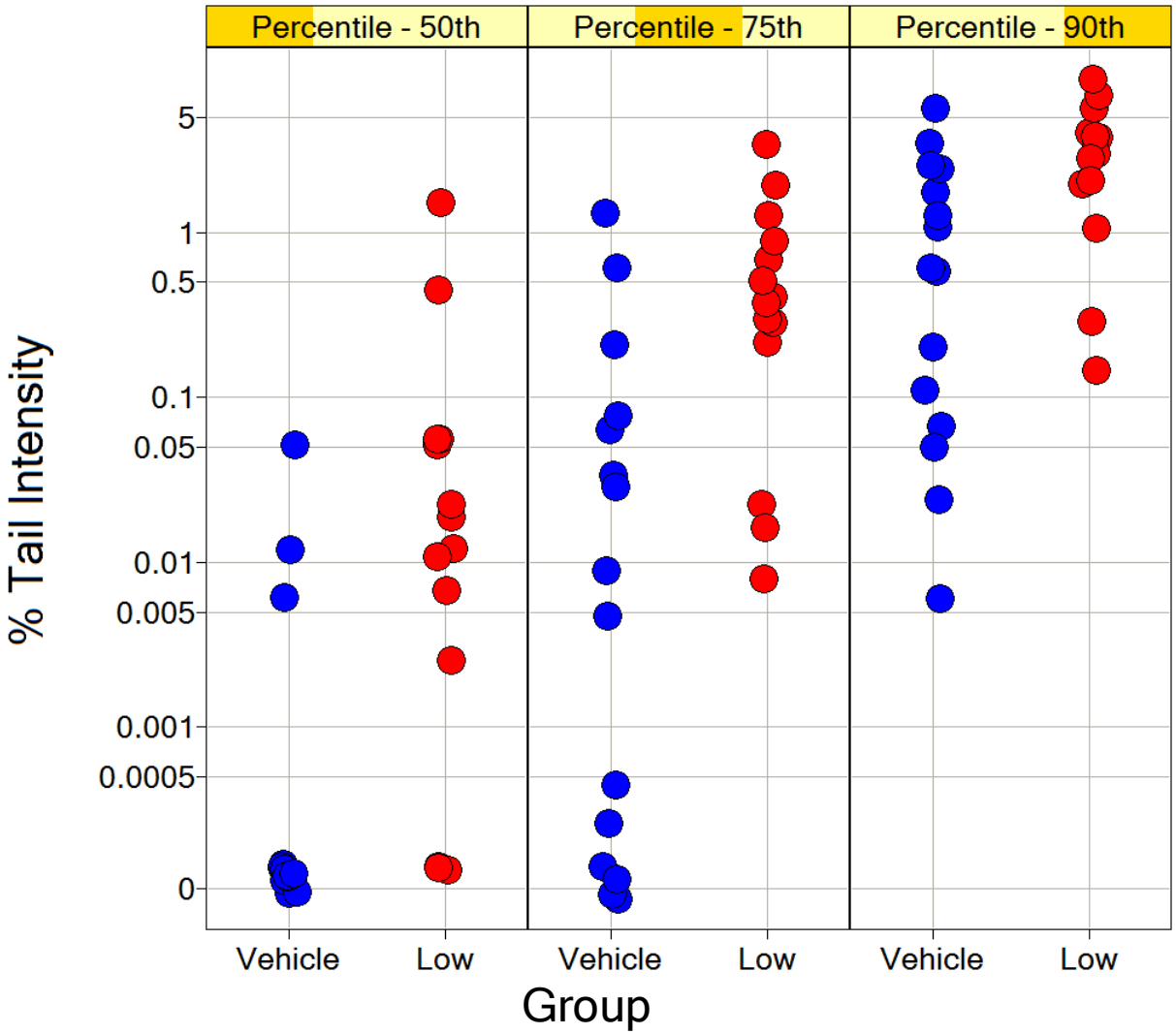
Summary Statistic 2. Example (linear scale)



Summary Statistic 2. Example (linear, jittered)



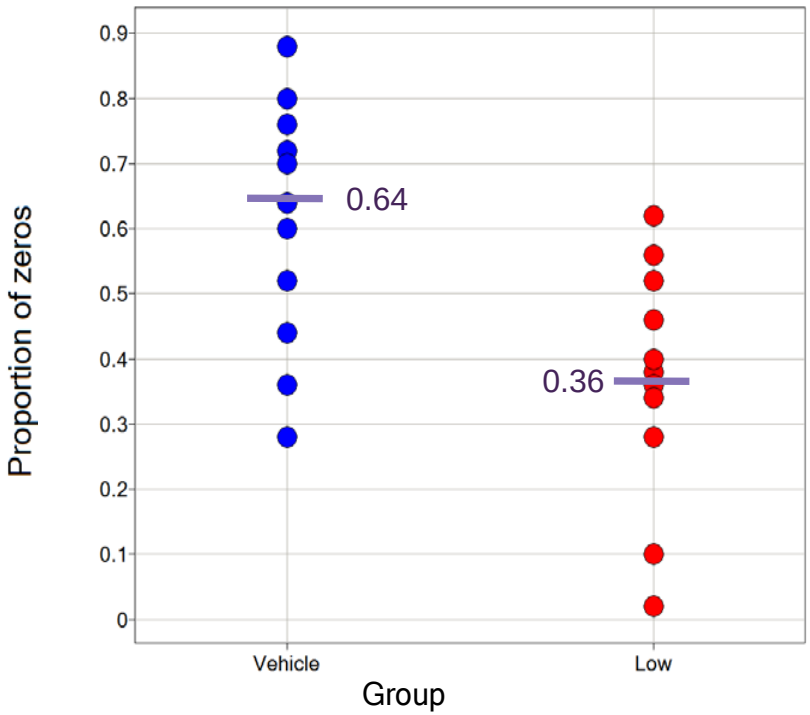
Summary Statistic 2. Example (log, jittered)

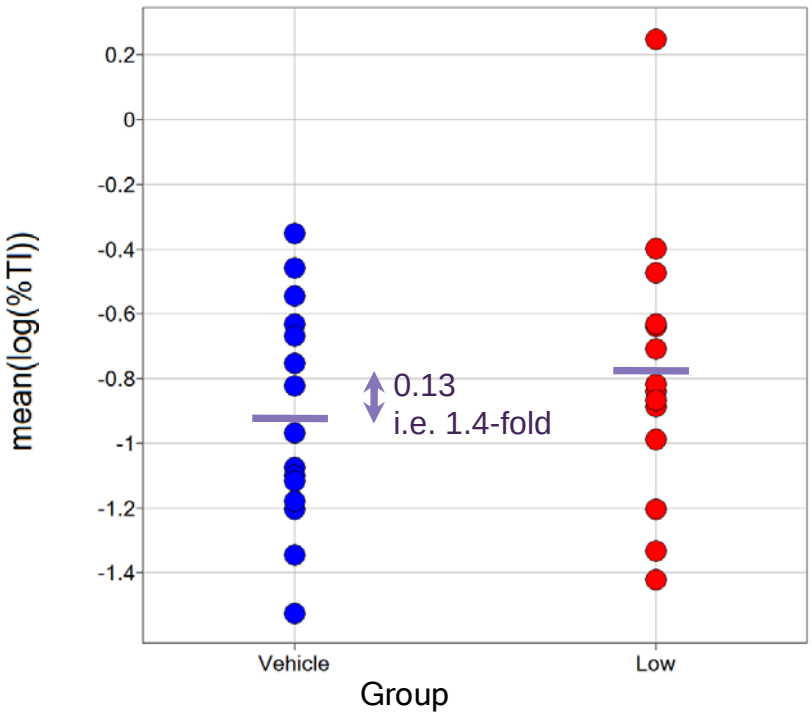


Two Summary Statistics?

- Awkward distribution of many zeros and small values plus some extreme values
- Too much to ask of a single summary statistic?
- Two summary statistics:
  - Proportion of zeros (or proportion of tail intensities < low threshold)
  - Mean(log(all other tail intensities)

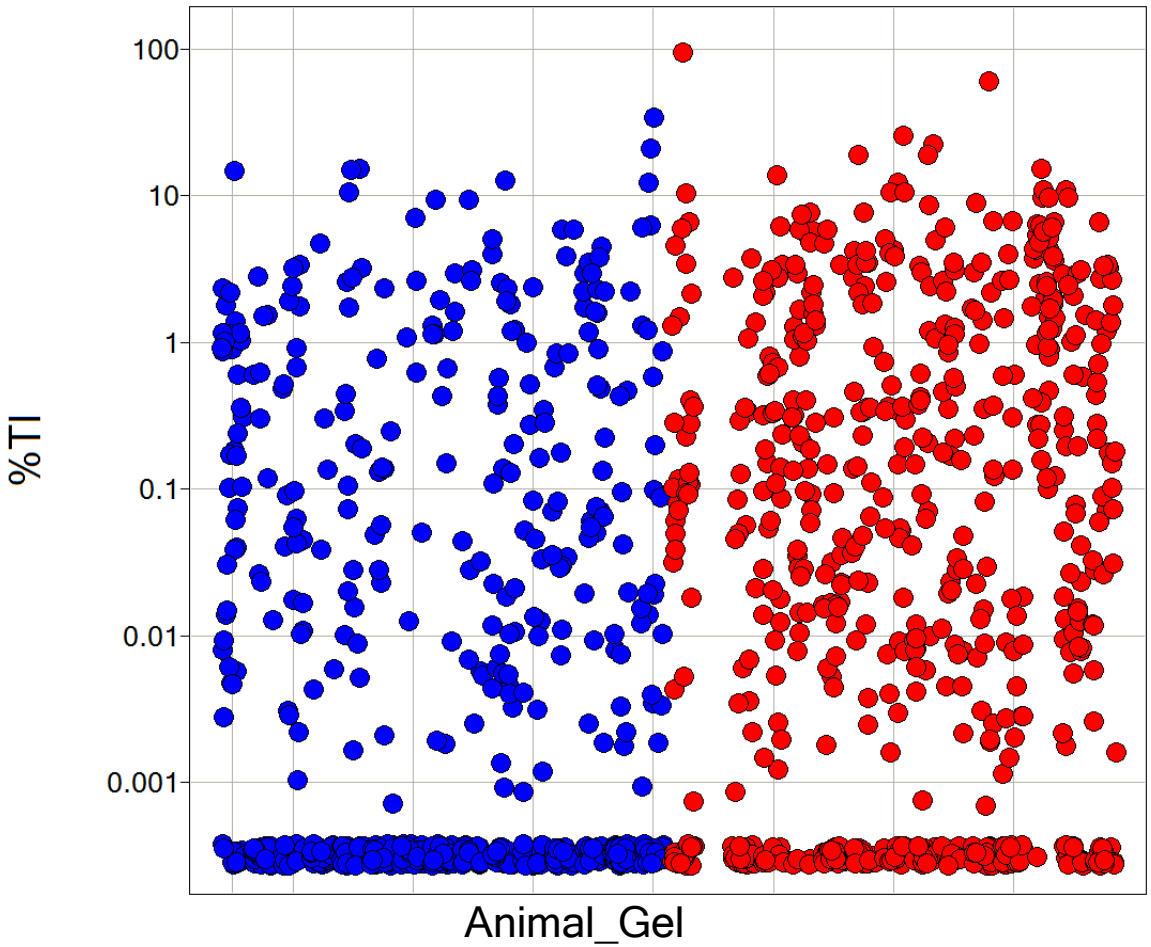
Two Summary Statistics. Example (linear, jittered)



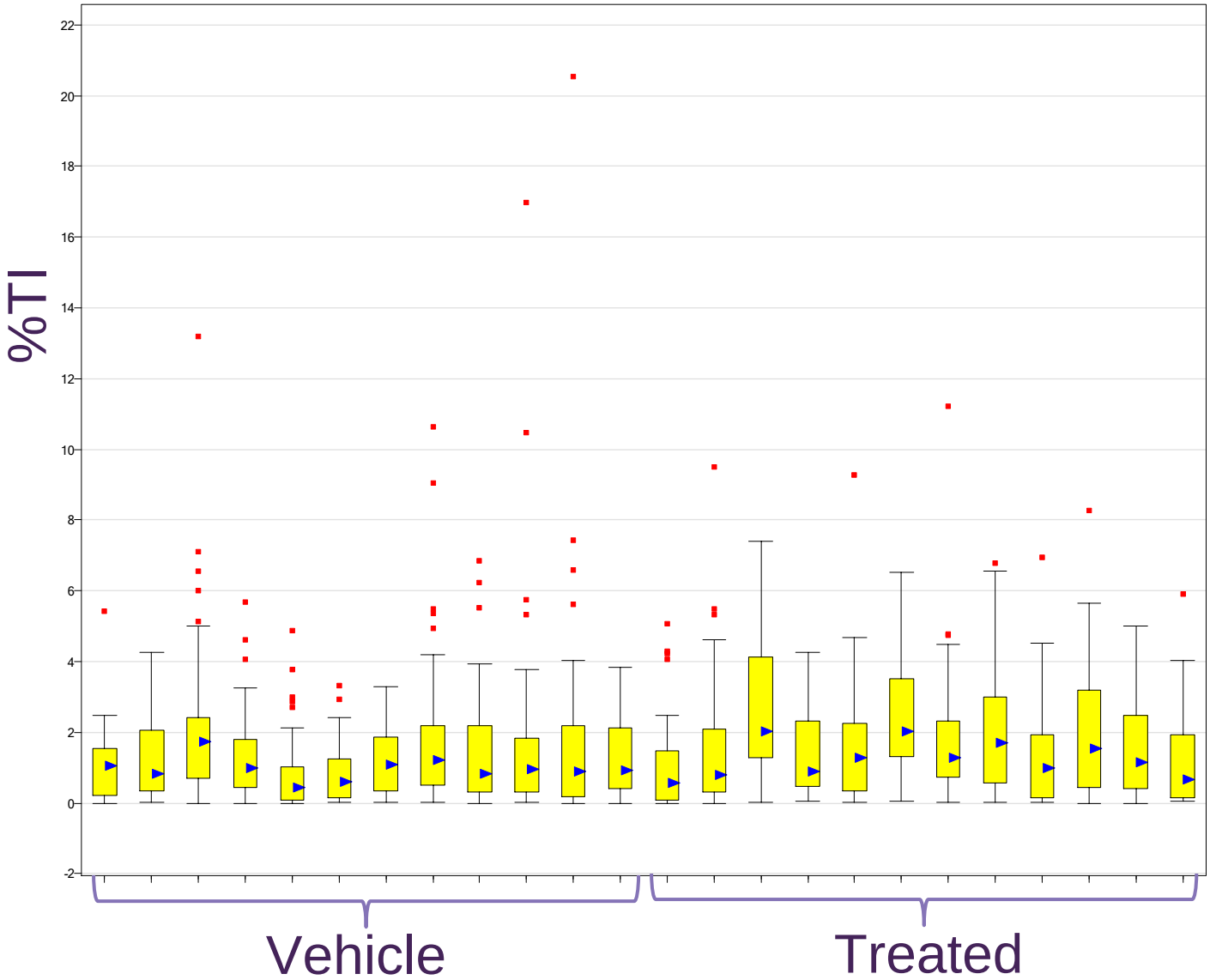


Recommendations

- Picture the raw data



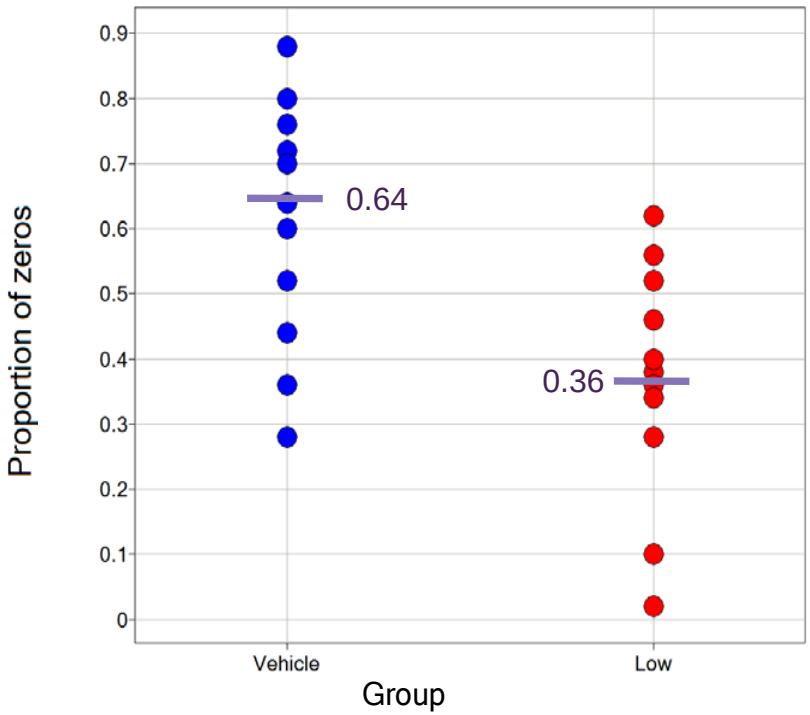
 Jonathan Bright, Comet Analysis – Issues, Limitations and Recommendations, NCSC 2010



Recommendations

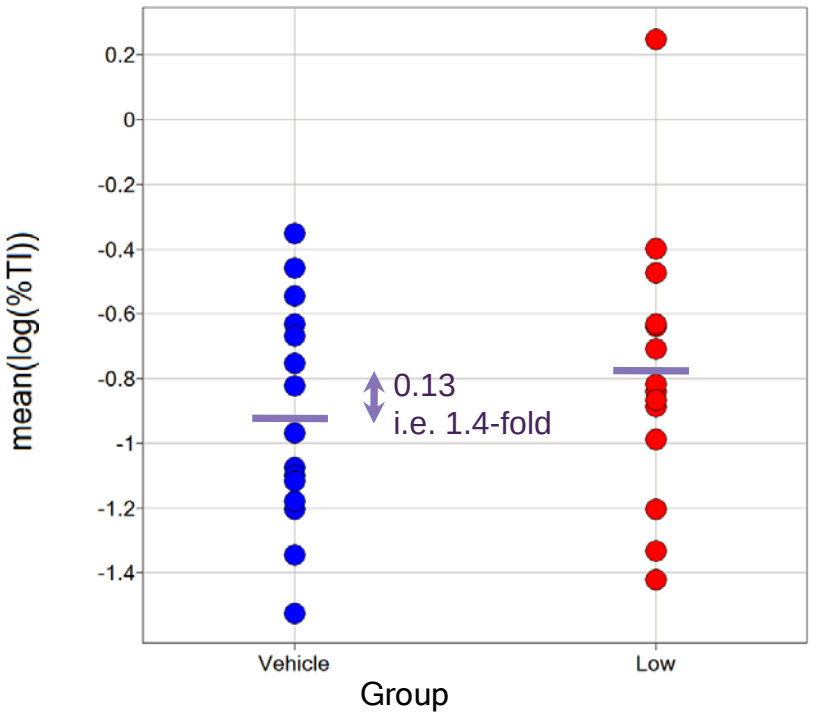
- Picture the raw data
- Consider using 2 summary statistics in the presence of awkward distributions
- Present results as confidence intervals

Two Summary Statistics. Example (linear, jittered)



95% 1-sided confidence interval extends up to 0.4

95% 2-sided confidence interval extends from 0.1 up to 0.45



95% 1-sided confidence interval extends up to 2.3-fold

95% 2-sided confidence interval extends from 0.7-fold up to 2.6-fold