

UX Statistics and Measures

Embedded Interface Design

with **Bruce Montgomery**

Learning Objectives

Students will be able to...

- Recognize different measures for usability assessment
- Consider methods in analysis of collected data
- Understand best practices for visualizing data and analysis



UX Statistics & Measures



- Occasionally there will be a desire to make a UX assessment more statistically significant
- This can especially come up in an engineering environment used to gathering data and measures for decisions or designs
 - How long does that operation take?
 - How many errors are people really making?
 - How do we know whether people like one design over another?
- Gathering, presenting, and making decisions from quantitative data is usually part of a formal usability or UX process

Typical UX Measures

- Task success counts – assess effectiveness of design
- Time on task – efficiency measure
- Error counts – efficiency and effectiveness
- Learnability – will the UX improve over time
- Satisfaction – often assessed by survey
- Mouse clicks, movement, or control activations - efficiency
- Problem/Issue counts – discovered during test, often weighted for severity
- Optimal path variance – how actions deviated from intended operation flow
- Reference [1]



Sample Size, Revisited

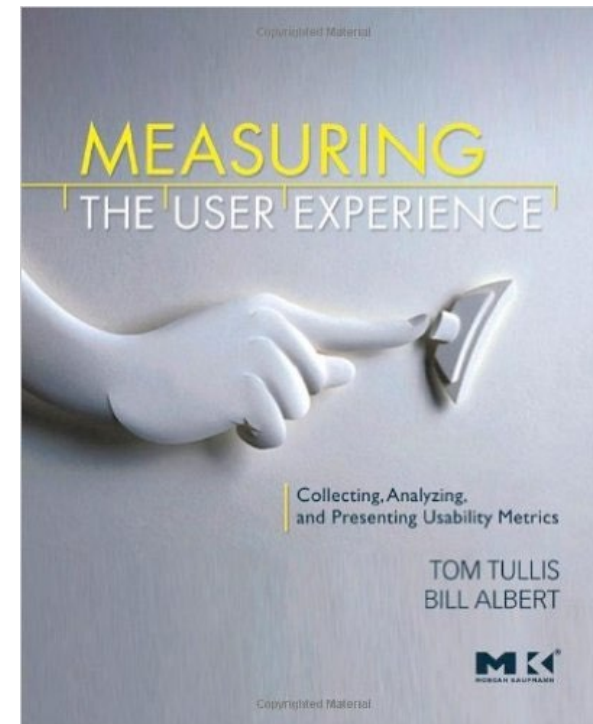
- A 2012 analysis of usability studies and the discovery of issues indicates that there is likely strong incompleteness for less than 30 participants, and that variance in issue visibility is a defining factor
- A prior referenced study suggested a sample of 10 +/- 2 users would find 80% of issues, but the more detailed analysis predicts a required sample size of 56!
- For qualitative work, iterating on small numbers of users makes sense, but quantitative work requires statistical sample sizing approaches [4]



Measuring UX - Tips

- Know your data
- Show your confidence (intervals)
- Deal with binary success data (appropriately)
- Compare means
- Consider using expectation measures
- Use the System Usability Scale (SUS)
- Show frequency distributions
- Combine different metrics
- Use appropriate tools
- Present data appropriately
- Reference [2]

- From Tullis & Albert's "Measuring the User Experience", an excellent book on metrics for usability [2]
- Great companion website at [5]
- Has papers, spreadsheets, tools, etc. – Highly recommended



Measuring UX - Myths

- Metrics take too much time to collect
- Usability metrics cost too much money
- Usability metrics are not useful when focusing on small improvements
- Usability metrics don't help us understand causes
- Usability data are too noisy
- You can just trust your gut
- Metrics don't apply to new products
- No metrics exist for the type of issues we are dealing with
- Metrics are not understood by management
- Difficult to collect reliable data with small sample size
- Reference [2]



Measuring UX – Statistics for Common Metrics

- Ordinal and Nominal vs. Intervals and Ratios
- Ordinal data = ranks, Nominal data = Categories
 - Ex: Poor/Fair/Good/Excellent, Y/N
- Analyzed as frequencies: Chi-square, Crosstabs, etc.
- Interval data = scale, Ratio data = times, averages
 - Ex: 1-7 Likert scale, SUS survey scores, average times
- Analyzed as means: descriptive statistics, t-tests, ANOVAs, correlation, regression analysis
- Reference [2]



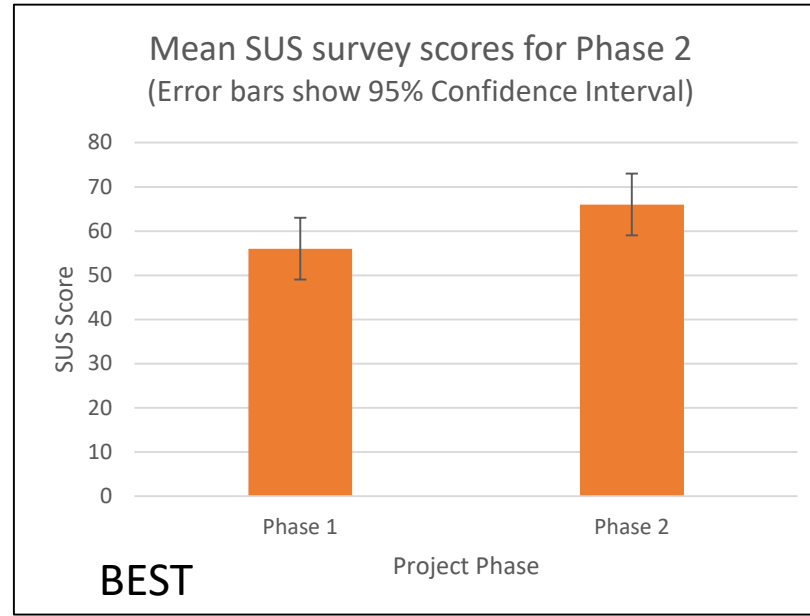
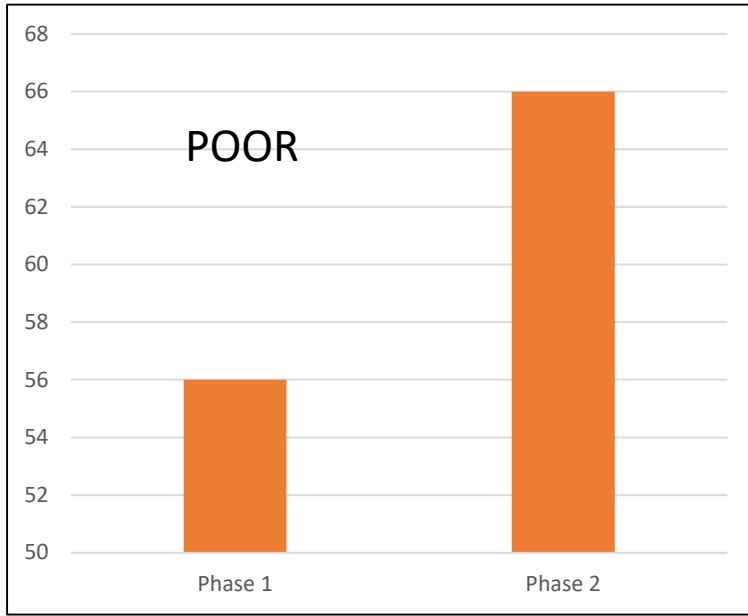
Measuring UX – Typical Data Visualization Errors

- Label axes and units
- Show confidence intervals when you can
- Start a bar graph at the lowest possible value for the y-axis (usually 0)
- Don't use a line graph when it should be a bar graph
- Poor labeling of pie charts
- Using color coding as the only way to convey information
- Use stacked bar graphs for showing parts of 100%
- Reference [2]

Tufte: It is key to be able to assess the credibility and integrity of data graphics. This includes documentation of data sources and the process used in data analysis. [3]



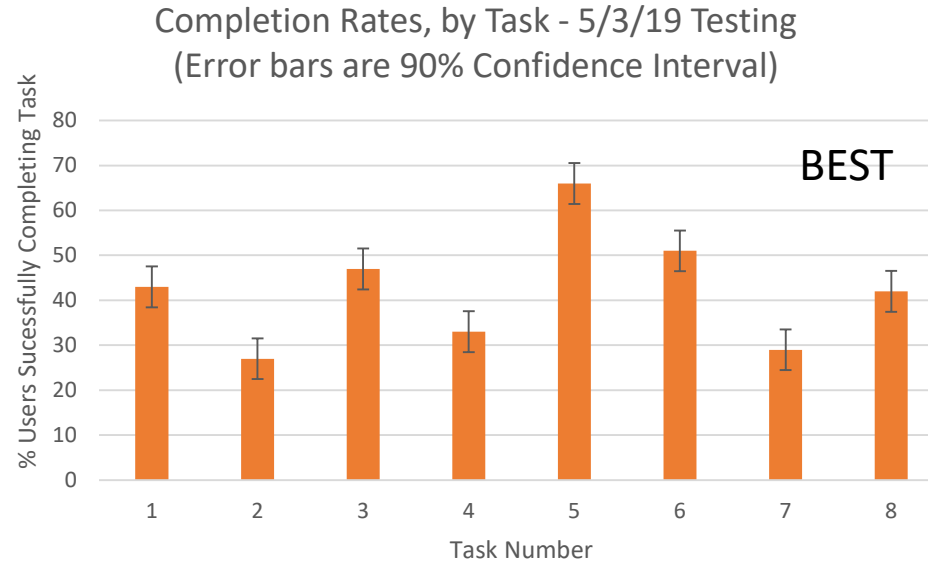
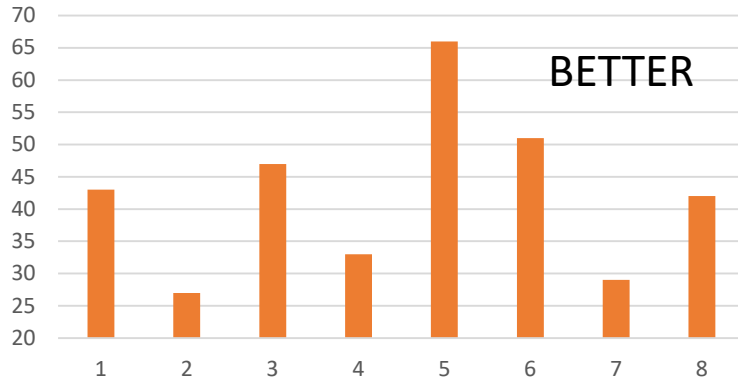
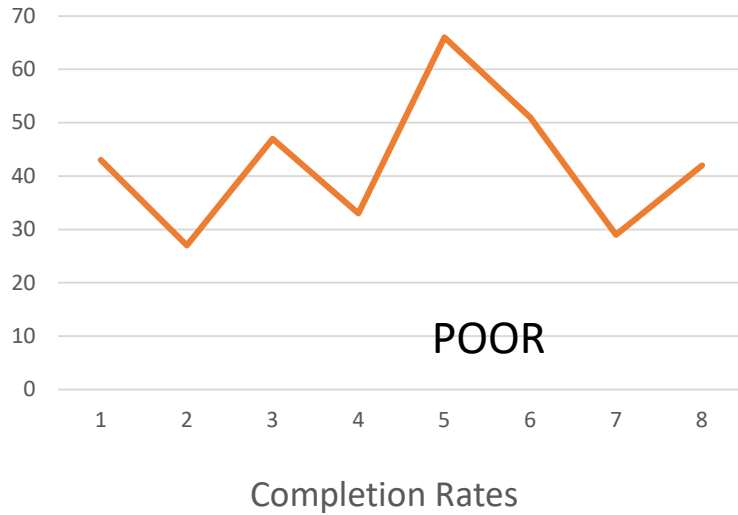
Measuring UX – Typical Data Visualization Errors



- First graph implies a significant difference in SUS scores
- Second graph shows confidence intervals overlap
- Same base data set...

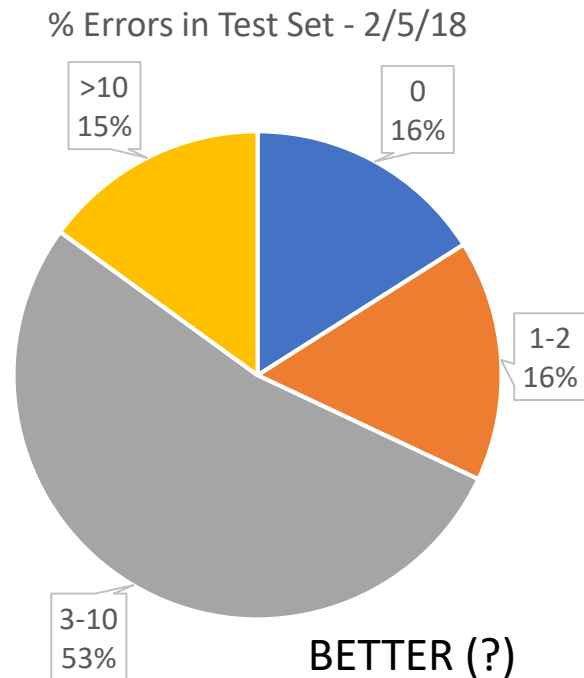
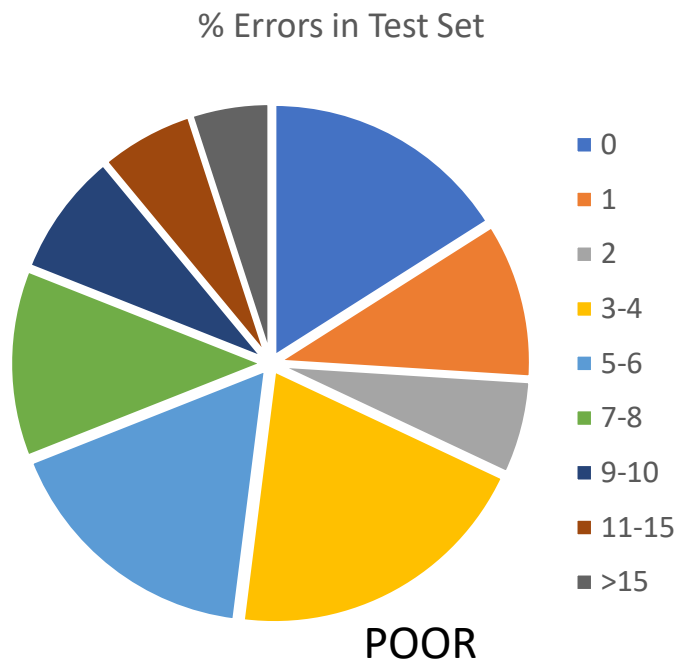
- First vs. second graph:
 - Missing titles, axis titles
 - Bar graph should start at 0
 - Missing CI error bars

Measuring UX – Typical Data Visualization Errors



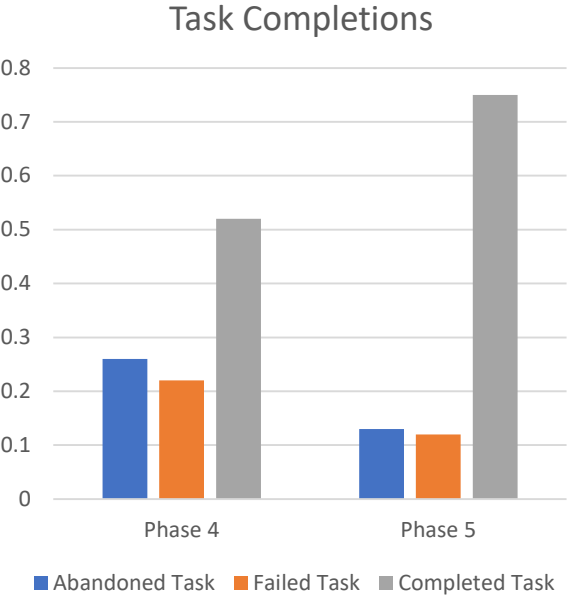
- Line graph implies a relationship, use a bar graph
- Bar graph should start at 0
- Include all labels and error bars if available

Measuring UX – Typical Data Visualization Errors

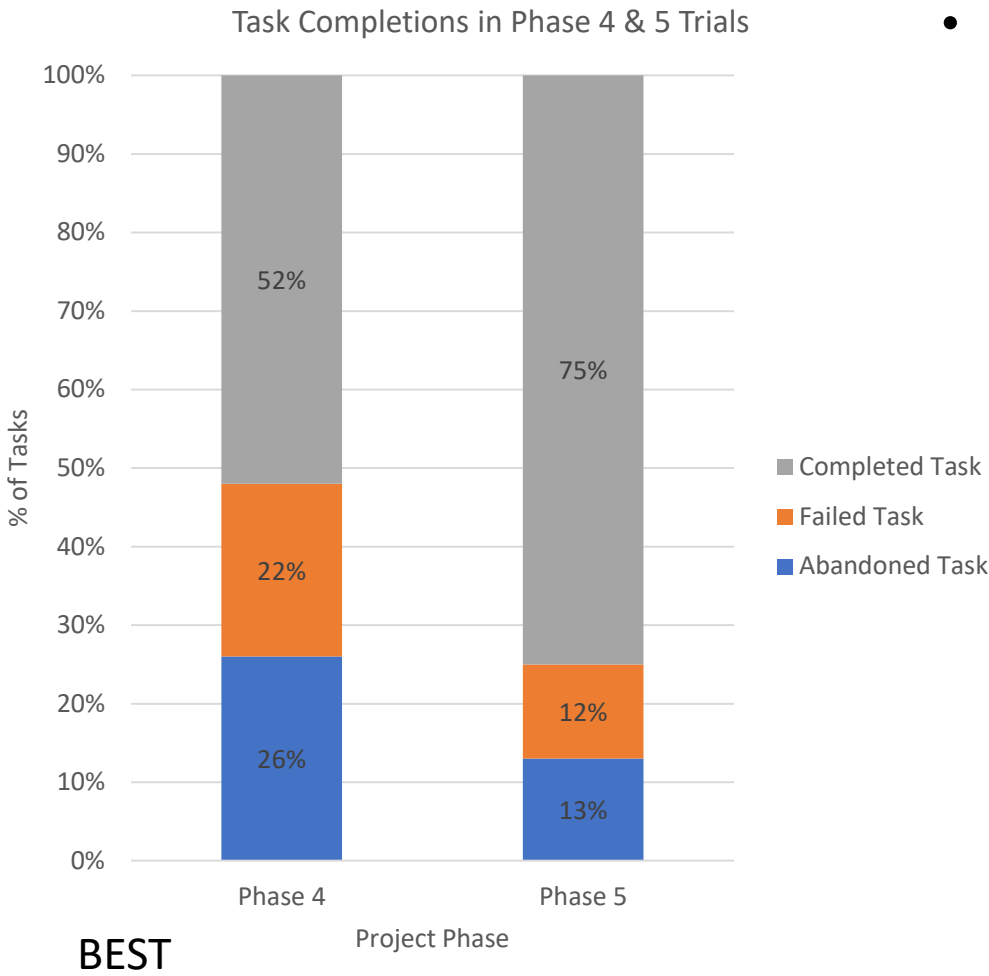


- Limit pie chart divisions to key categories
- Do not use color as the only identifier
- Montgomery – generally avoid pie charts for better options

Measuring UX – Typical Data Visualization Errors



POOR

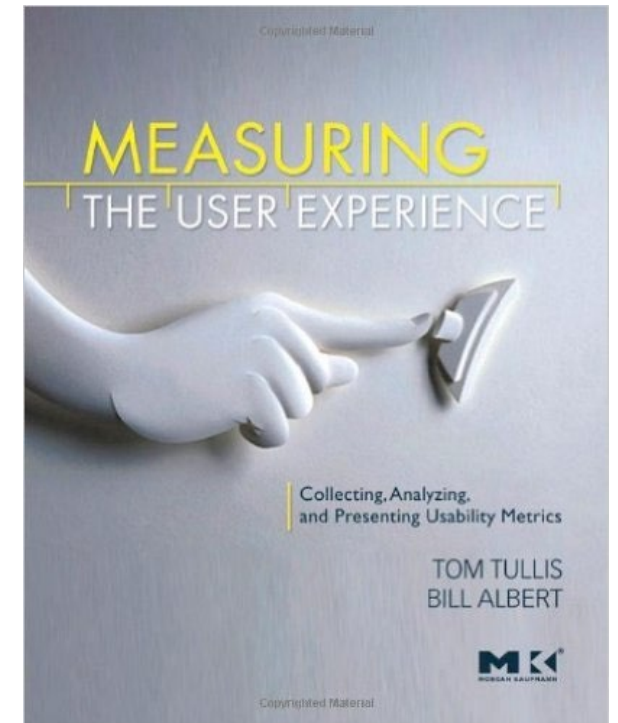


BEST

- If measures add to 100%, stacked bars can provide a better comparison than pies or other bar charts

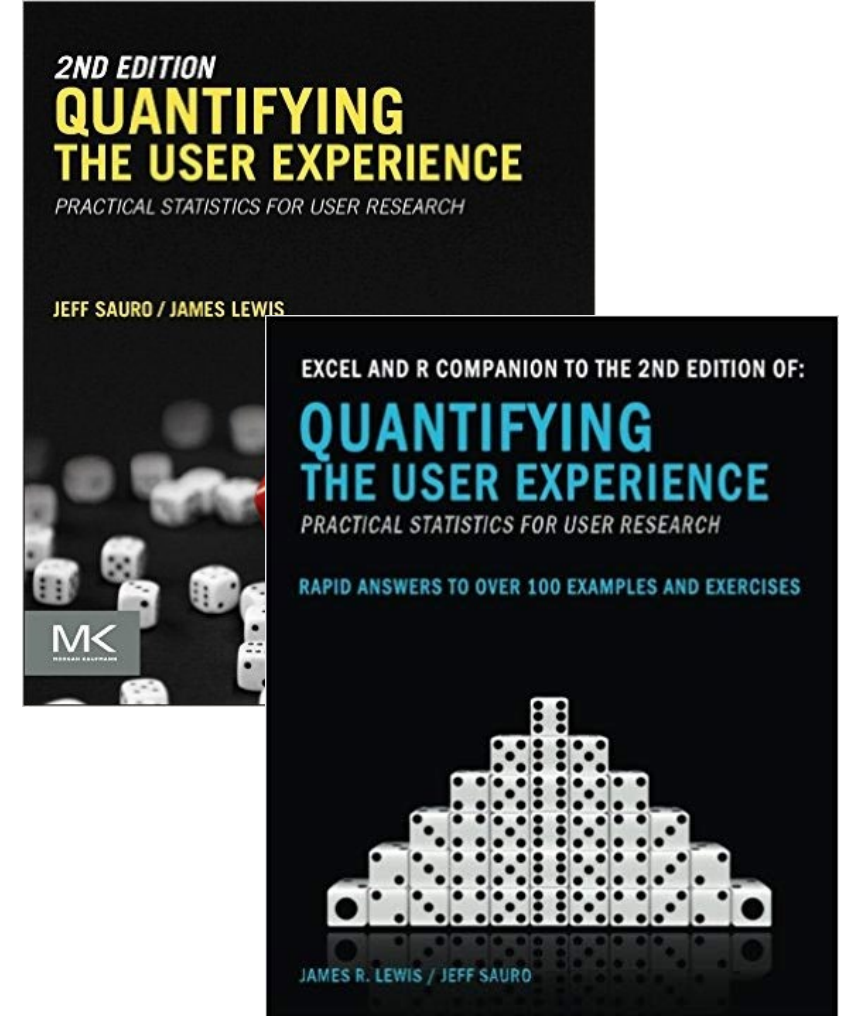
Recommended Resources

- Tullis & Albert's "Measuring the User Experience" is an excellent book on metrics for usability [2]
 - Great companion website at [5]
 - Has papers, spreadsheets, tools, etc. – Highly recommended



Additional Resources

- “Quantifying the User Experience” by Sauro & Lewis is another standard text on statistics for UX and usability [6]
- Key elements:
 - Summarizing data and margins of error
 - Determining statistical significance
 - Finding appropriate sample sizes
 - Comprehensive overview of standardized usability questionnaires
- Companion book has Excel and R routines for standard UX related analysis [7]
- Principals have a web site [8]
- Similar articles/blogs on related topics [9], [10]



Summary

- Of course this can be a fairly deep topic if you get into the details of the statistical analysis methods, for that I highly recommend the resources identified and their corresponding web support
- But using a standard survey, like the SUS, and then displaying before and after results well, can be a compelling statement on usability improvement, and it's fairly easy to do
- Remember though, use the methods appropriate to your goals, you may not need this level of rigor to make a UX design better



References

- [1] <http://webservices.itcs.umich.edu/drupal/wwwsig/sites/webservices.itcs.umich.edu.drupal.wwwsig/files/Usability-Testing-Basics.pdf>
- [2] Measuring the User Experience, Tullis & Albert, 2013, Morgan Kaufmann
- [3] https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003wa
- [4] https://www.researchgate.net/publication/233842847_Sample_Size_in_Usability_Studies
- [5] <http://www.measuringux.com/>
- [6] Quantifying the User Experience, Sauro & Lewis, 2016, Morgan Kaufmann
- [7] Excel and R Companion to Quantifying the User Experience, Sauro & Lewis, 2016, Createspace
- [8] <https://measuringu.com/>
- [9] <https://measuringu.com/graphing-displaying-data/>
- [10] <https://measuringu.com/five-hard-quant/>

