

Evaluation plan

Scenario

Users: Scottish government (environmental/local planning department)

Environment: at work, on each persons computer (or a print out if they prefer)

5 tasks

1. Which Local Authority had the 3rd largest percentage point change in % of adults close to green or blue space between 2013 and 2018?

Answer: Dundee City (-21%)

Testing: Ability to understand ranking of differences.

2. How many LAs had positive, negative and no change between 2013 and 2018?

Answer: Positive 14, Unchanged 2, Negative 16

Testing: Ability to distinguish between positive, none and negative changes.

3. What was the difference in the percentage point change in % of adults close to green or blue space between 2013 and 2018 between North Ayrshire and Midlothian?

Answer: 2%

Testing: Ability to compare differences in LAs with both the same sign of change (here both positive changes).

4. Which LA had the highest % of adults close to green or blue space in 2018?

Answer: Shetland Islands

Testing: Ability to read and compare values at 2018.

5. Which LA had the highest % of adults close to green or blue space in 2013?

Answer: Orkney Islands

Testing: Ability to read and compare values at 2013.

What are you measuring?

These test whether the following aspects, which (as logged in journal) need to communicate through the visual are clear:

- Values at the 2 time points (2013 and 2018)
- Direction and size of change between time points
- Easy and precise comparisons between LAs

What study design, protocols or methodology do you want to follow?

One-off, target user reviews using heuristics

What data are you collecting?

Quantitative

- Time taken to answer each question
- Answers to questions

Qualitative

- If unable to answer question leave open to any explanation of what found unclear
- There was something the visual caused them to want to know more about.

Link to Google Form questionnaires

Pre-evaluation: <https://forms.gle/5oYpG8vBdezUiXTx5>

Post-evaluation: <https://forms.gle/rkoXinFPQ2X17Jj57>

How will you analyse or evaluate your results?

Time metric

Compare against some benchmark. Perhaps comparing against time taken to answer the same task questions but seeing this data presented as bar chart (as common chart type people may have familiarity with) or if there is a benchmark from research that is used as a 'acceptable' time takes for someone to analyse a visual.

Error metric

Have a benchmark predefined prior to performing the evaluation that see as 'acceptable' margin of error (this benchmark may differ depending on the backgrounds of the users based on the pre-evaluation survey e.g. if they are frequent visualisation users may have a higher benchmark than if users tests use visualisations infrequently).

Qualitative

Look to see if common themes/comments are occurring from users and can consider taking these on board (weighing this up will 'over-evaluating' and not being able to fit a specific need/feedback of specific user).