Analytical Literacy Summary

The Analysis Function

▲ The Analysis Function is a cross-government network bringing together a range of analysts including Statisticians, Data Scientists, Economists etc. There are many different roles within the Analysis Function, each of which who produce different types of analysis.

Asking analytical questions

▲ The Analysis Lifecycle describes the iterative process whereby analysis is scoped, designed, collected, and delivered. This should be refined throughout the process as new information is uncovered.

▲ The entire analysis process should be collaborative between users and analysts to ensure requirements are balanced with resources and capabilities. This is particularly true during the scoping and design stages.

▲ Key requirements of analytical work are that it should be repeatable, independent of bias, relevant to the questions, objective and robust enough to survive challenges.

Methods of analysis

▲ **Sample**: a sample is a smaller subset of the entire population, usually aiming to represent data of the entire population with less resources.

▲ **Weighting**: attaching a weight to each observation can correct imbalances in proportions of certain characteristics to make a sample more representative of the population.

▲ **Statistical Significance**: a statistically significant result is one unlikely to have occurred due to change and is therefore likely to represent the true figure in the population.

▲ **Confidence intervals**: a range of values that the true value of the population is estimated to fall between.

▲ **Confidence level**: the probability that the true value will fall between the given confidence intervals.

▲ **Regression**: a kind of analysis which quantifies the strength and nature of a relationship between two variables.

▲ **Relative risk**: the likelihood of a given event occurring, in relation to another event, rather than the probability of it occurring overall.

▲ **Absolute risk**: The chance/probability of a given event happening, often in percentage form.

▲ **Time series**: a set of time ordered data, usually collected at regular intervals such as daily or monthly, such as the Consumer Prices Index.

Analytical Literacy Summary

▲ **Data Science**: an inter-disciplinary field, bringing together mathematics, coding skills and making the most of new innovative data tools.

▲ **Data linkage**: the process of trying to establish whether two records from two different databases relate to the same entity

Quality assurance

▲ With analysis being data-driven, the quality of data is essential, as any issues with the data can cause problems throughout the rest of the analytical lifecycle.

▲ Producing high quality statistics is fundamental to building public trust. This is done by using suitable data, sound methods and assuring with details of production.

▲ Reproducible analytical pipelines (RAP) are automated analytical process, usually using openly accessible software that incorporate best practices and reproducibility throughout the analytical process.

Data Visualisation

▲ Data visualisations are useful for exploring data or to communicate messages. They can make data much more digestible than through a block of text. All visualisations should follow the best practice guidelines.

▲ However, data visualisations can be misleading and users should ensure they think critically when drawing insights from visualisations. One method that may produce misleading visualisations is cherry-picking; where a subset of data is used that only tells part of the story.

Drawing and communicating insight

▲ Analytical outputs support decision-makers in making robust, evidence-based decisions. It is important for this process to be collaborative throughout, between analysts and users to ensure expectations are achievable based on resource and explanatory power.

▲ Using a range of analytical outputs is the best way to ensure the evidence-base is strong and accurate.

▲ Phrasing and graphics should be tailored to the message and in as plain English as possible. It is also essential to ensure that uncertainty is conveyed in the results.