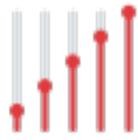# Measuring Broadband America

Data Processing Flow, Eleventh Report

The following diagram describes how raw data was processed for the production of the Measuring Broadband America Eleventh Report:

Measuring Broadband America

Data Processing Flow Steps:

1 Raw Data
Raw data for the chosen period is collected from the measurement database. The ISPs and products that panellists were on are exported to a 'Unit Profile' file. ISP and product assignments are obtained through the validation process operated in conjunction with ISPs.
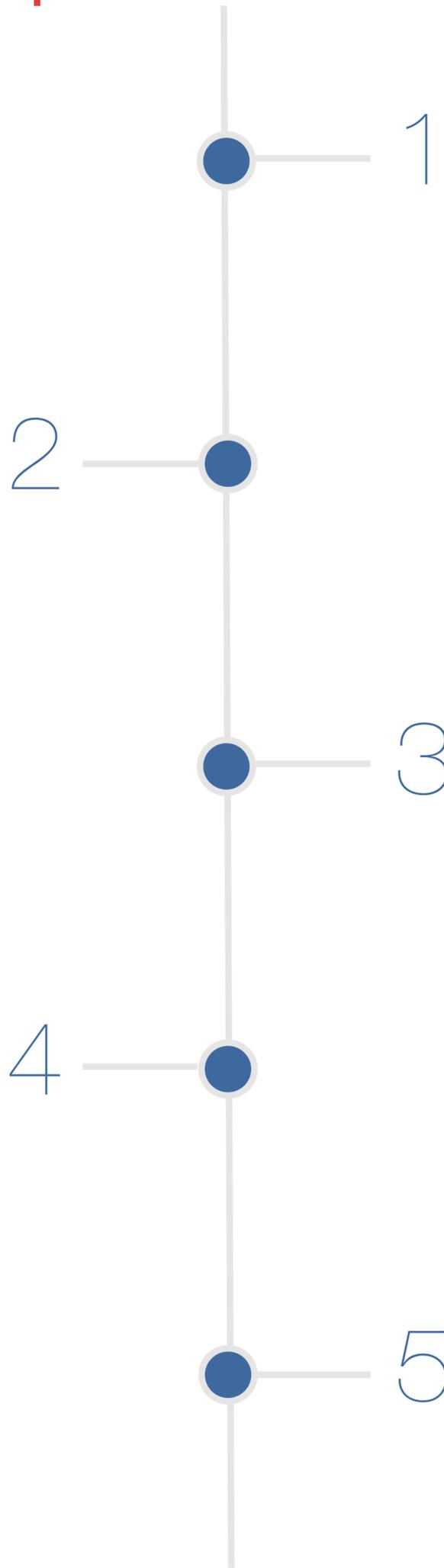
2 Clean Data
Data is cleaned. This includes removing measurements when a user changed ISP/product during the period. Anomalies and significant outliers are also removed at this point. SQL scripts used to clean the data are published under the name 'SQL Cleanup Scripts'.

3 Per-device Aggregates
Per-device aggregates are generated for each metric. It is at this point that the 7-11pm averages are computed, along with trimmed means, medians and various percentiles for each metric. The R script used is published under the name 'R Processing Script'.

4 R Processing
The per-device, per-metric CSV data is aggregated further and coupled with the unit profile data. ISPs and products with low sample sizes are excluded at this point.

5 Excel Tables / Charts
Aggregated data is transformed into an Excel workbook, which is published as 'Statistical Averages'.