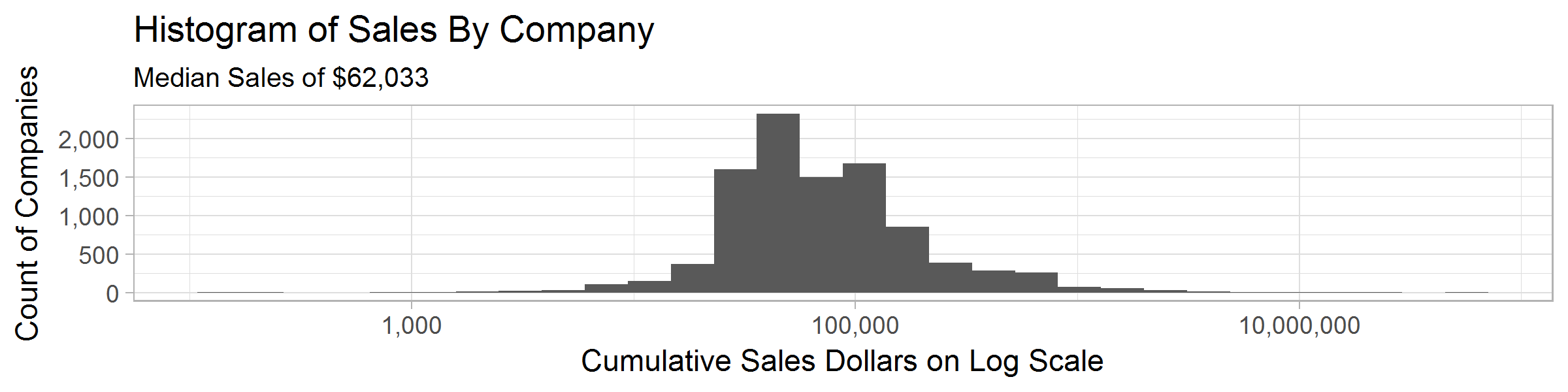
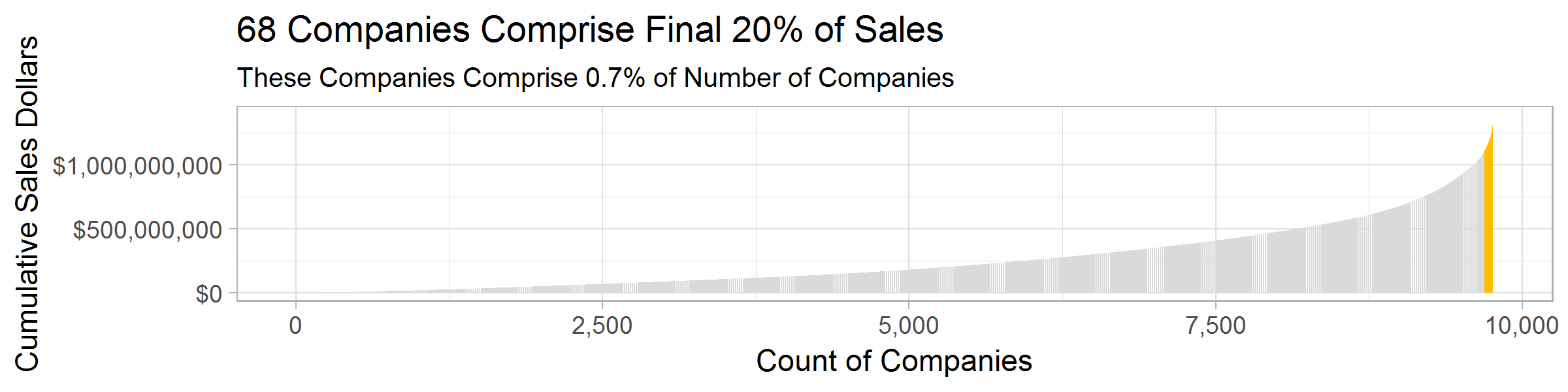
Certainty Stratum – Methodology and Selection

Exploration of the data found useful information about the shape of the distribution of the data set for Sales by Company. Among the 9,762 companies included, the Mean Sales was $142,267, but with a much lower median - $62,033. This indicates a long tail distribution – with several outliers on the upper end of sales that are likely increasing the mean as well as the standard deviation of the sales value.



Considering this information, we investigated the possibility that a small number of the companies comprises a large share of the overall sales. In any scenario where the variables of interest – in this case, the total sales of the population – are of interest, a certainty stratum that comprises a large share of the overall value might allow a better insight into the final result. In a non-probability study, this could be done as a cut-off sample – but with the proposed survey design for this study, we propose that the companies providing the final 20% of cumulative sales to be taken as a certainty stratum.



This comprises 68 companies, but which sell 20% of the overall value of all companies in the population. This will allow for 432 of the 500 samples to be allocated to the remainder of the population, 9,693 companies. This gives an overall probability of selection for the remaining companies of 0.045. Overall probability of selection is 0.051 (500/9,762).

Stratum Creation and Break Points

With the certainty stratum determined, the next consideration is the creation of the stratum and the allocation of sampling units by stratum.

Two methods of the

Determination of Stratum