Applied Data Analytics

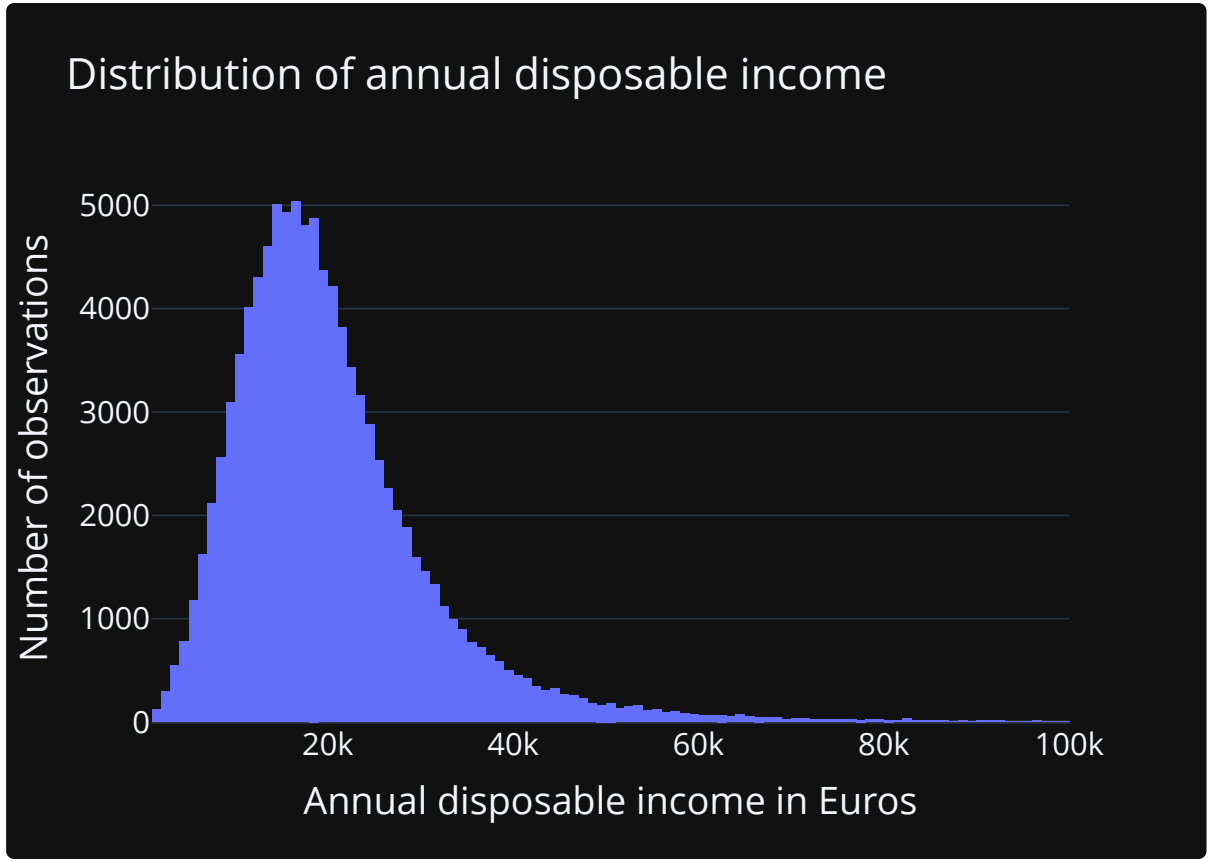
Statistics — Basics & location

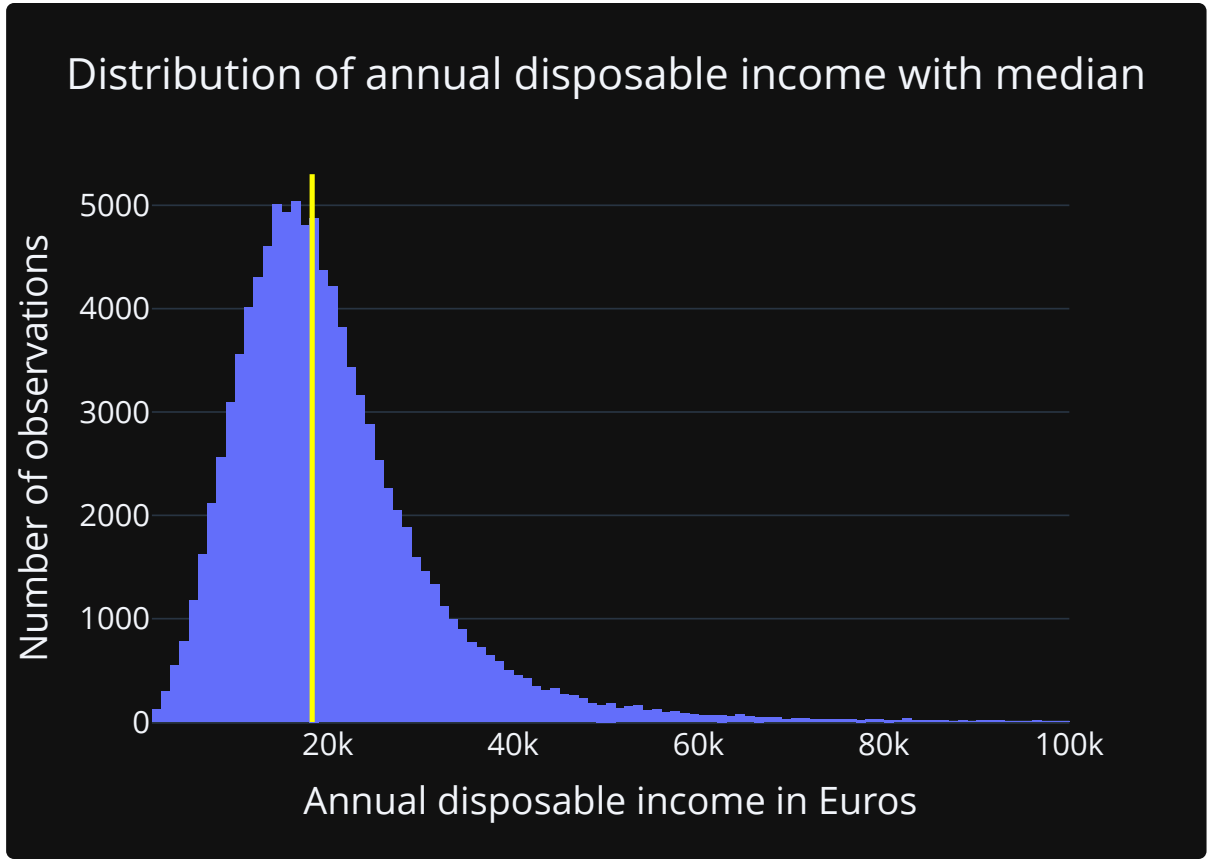
Quantiles

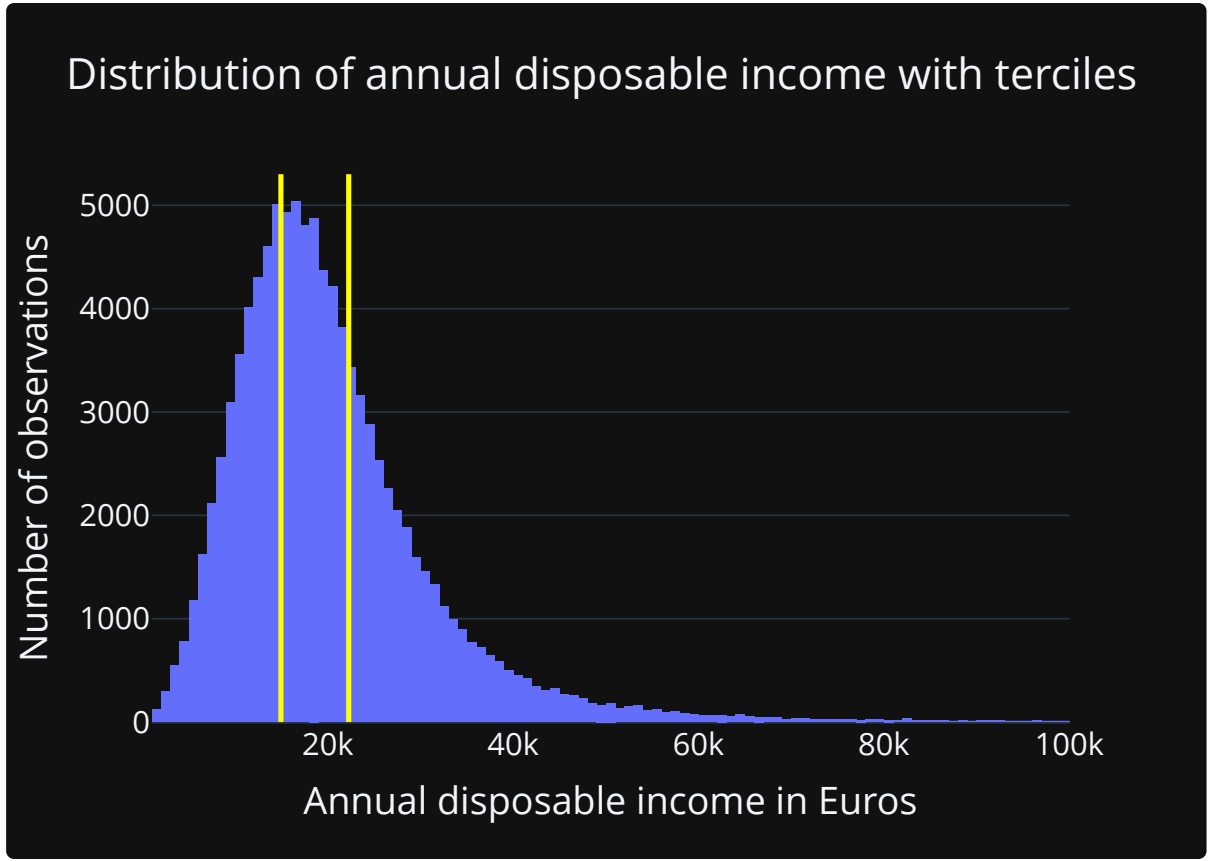
Hans-Martin von Gaudecker

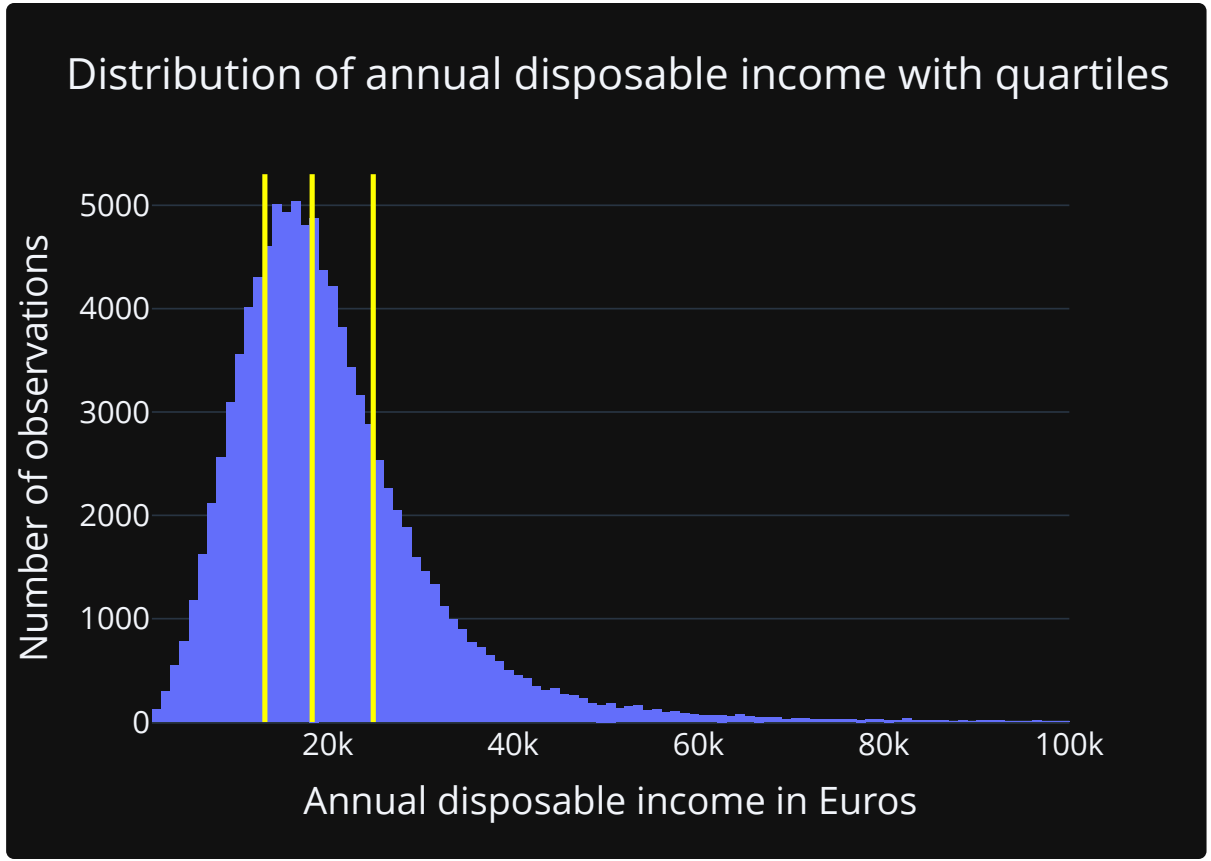
- The median marks the halfway point of a distribution.
- Terciles, quartiles and quintiles use cut points at thirds, quarters and fifths.
- Deciles and percentiles use cut points at tenths and hundredths.

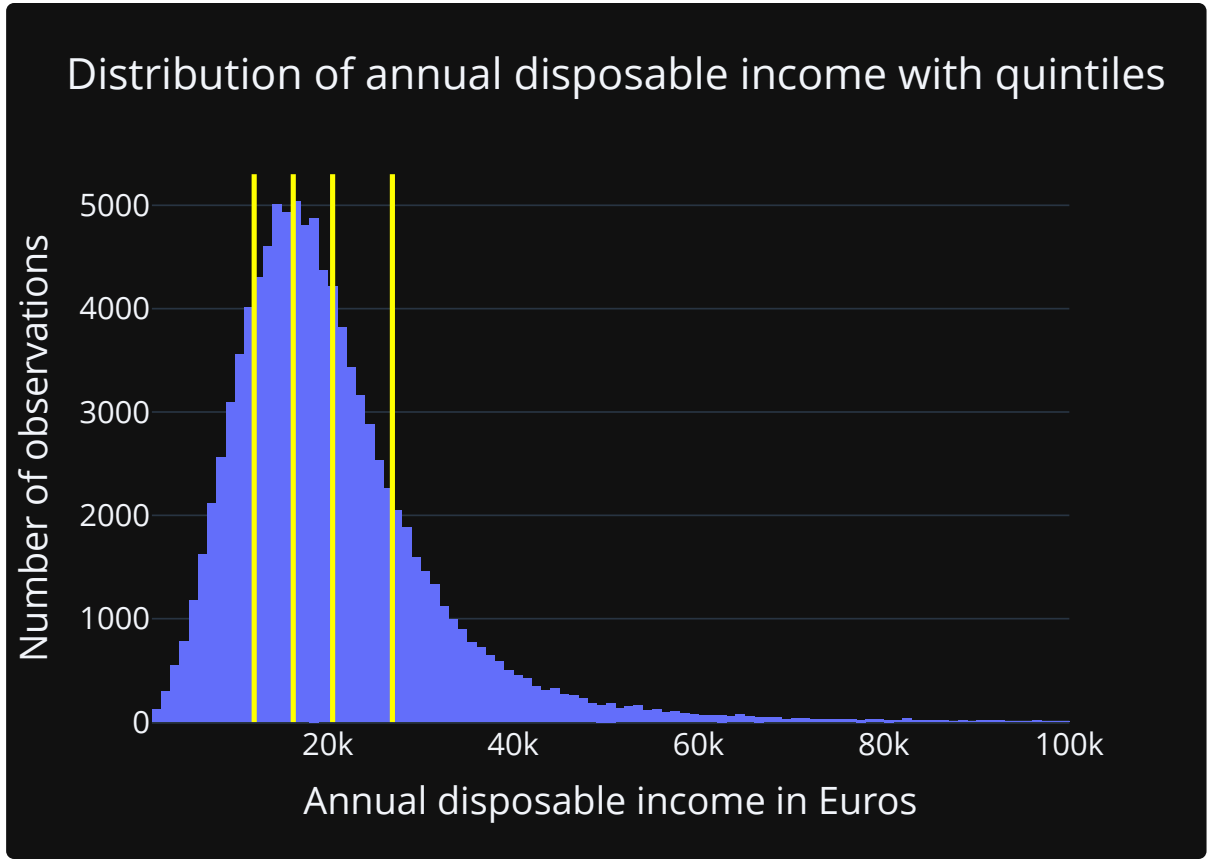
These cut points give equally probable bins for continuous distributions. In finite samples, ties and indivisible observations can prevent exactly equal bin counts.

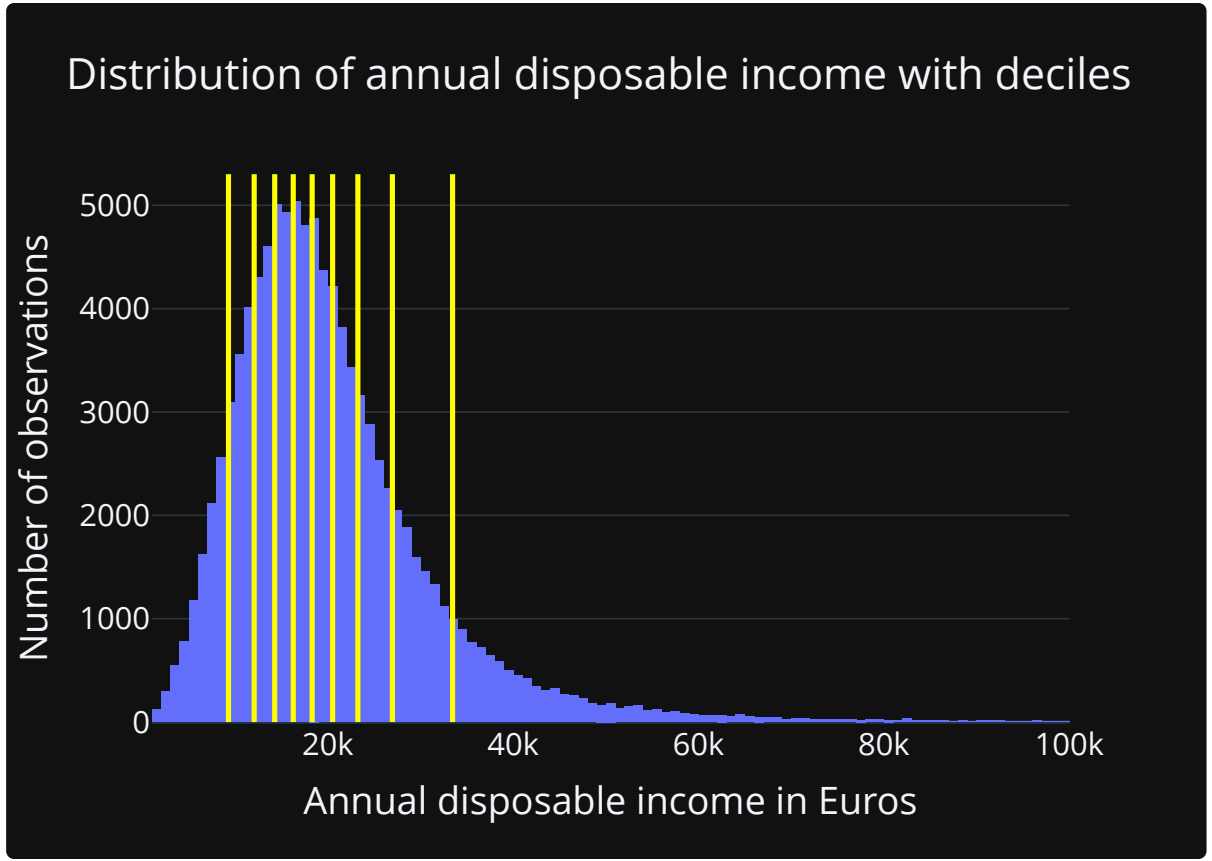

Distribution of annual disposable income
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros

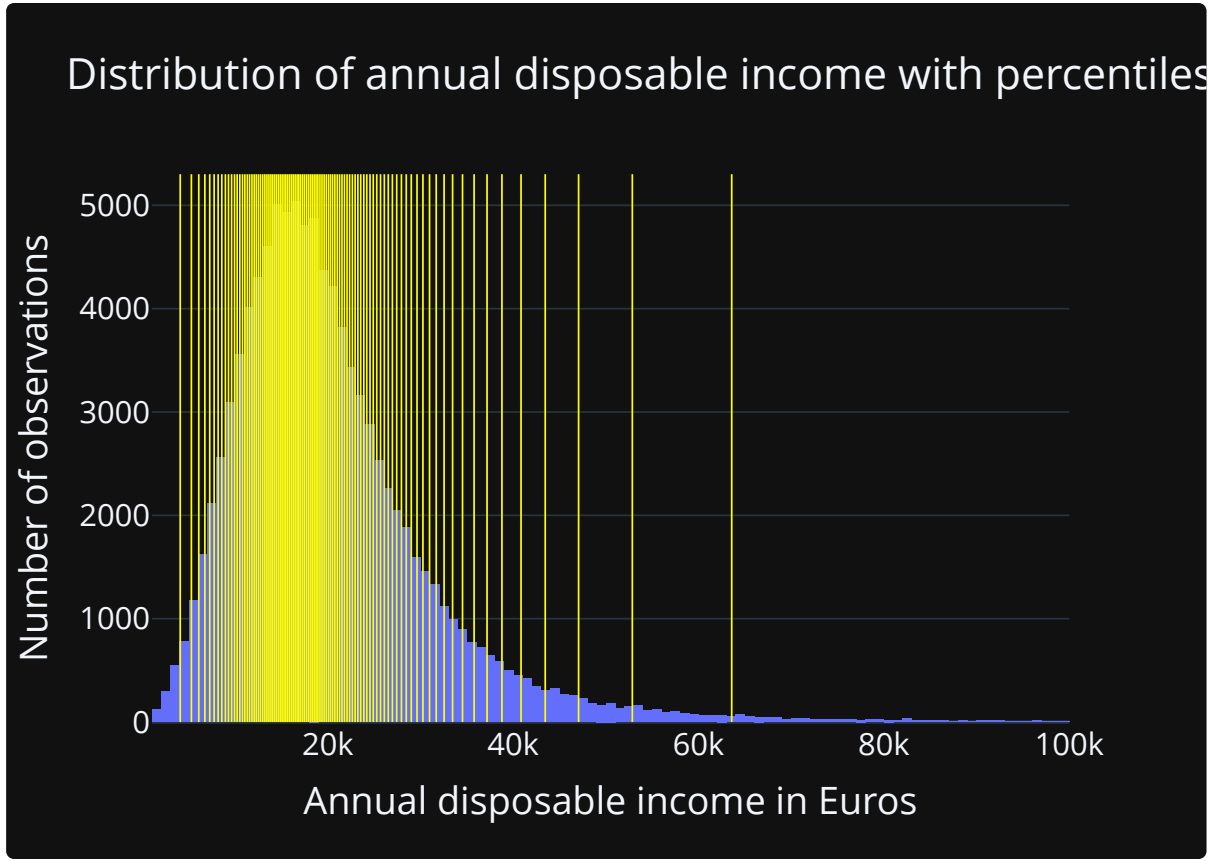

Distribution of annual disposable income with median
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


Distribution of annual disposable income with terciles
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


Distribution of annual disposable income with quartiles
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


Distribution of annual disposable income with quintiles
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


Distribution of annual disposable income with deciles
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


Distribution of annual disposable income with percentiles
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros

$$\in [0, 1]$$

$0 < q < 1$ q x

q x

$1 - q$ x

$q = 0$ $q = 1$

Terminology often a bit loose

We often refer to things like "the top quartile", meaning all observations above the 3rd quartile / 0.75-quantile.

To avoid ambiguous phrases such as "in the 90th percentile", specify the bin explicitly. For example, a 90%–91% quantile bin contains observations such that

$$q_{0.9} \leq x_i < q_{0.91}$$

This refers to a bin instead of a point; with ties it may be empty or have more than 1% of observations.