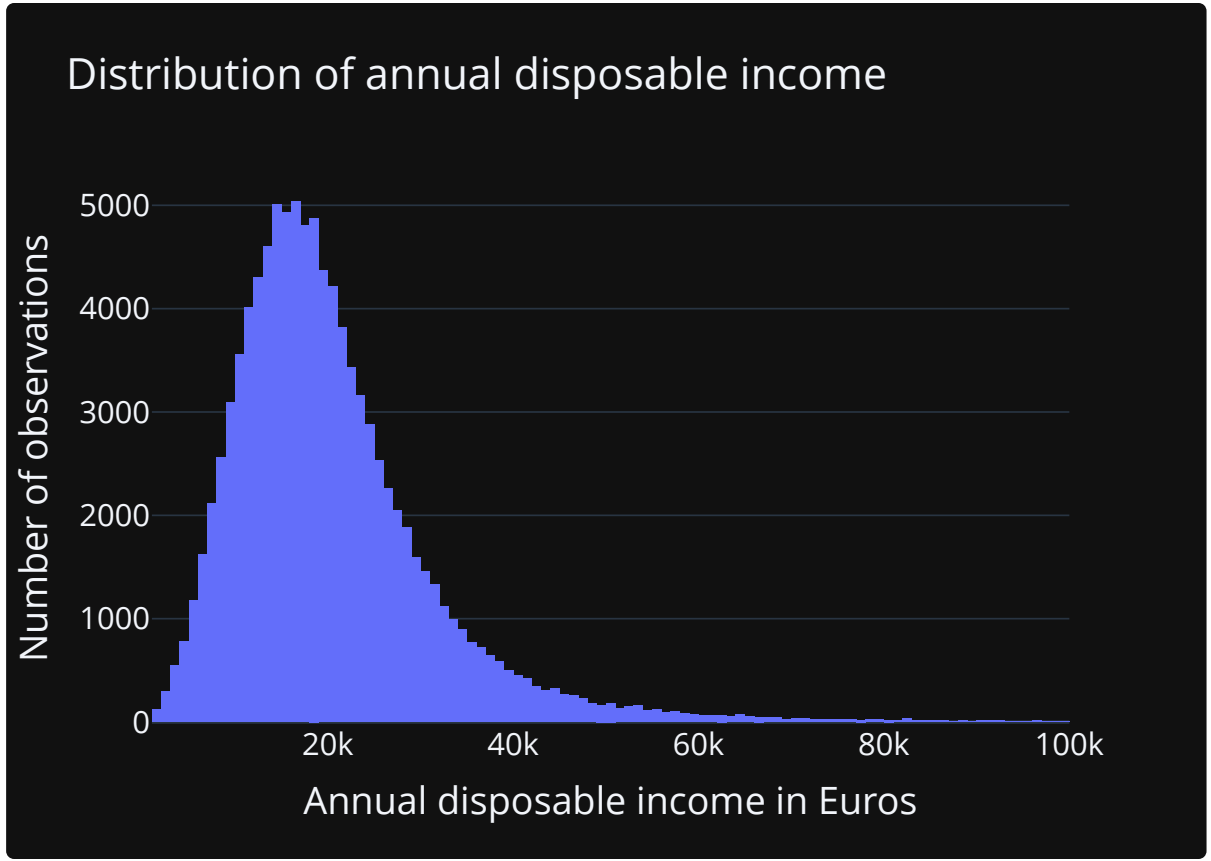
Applied Data Analytics

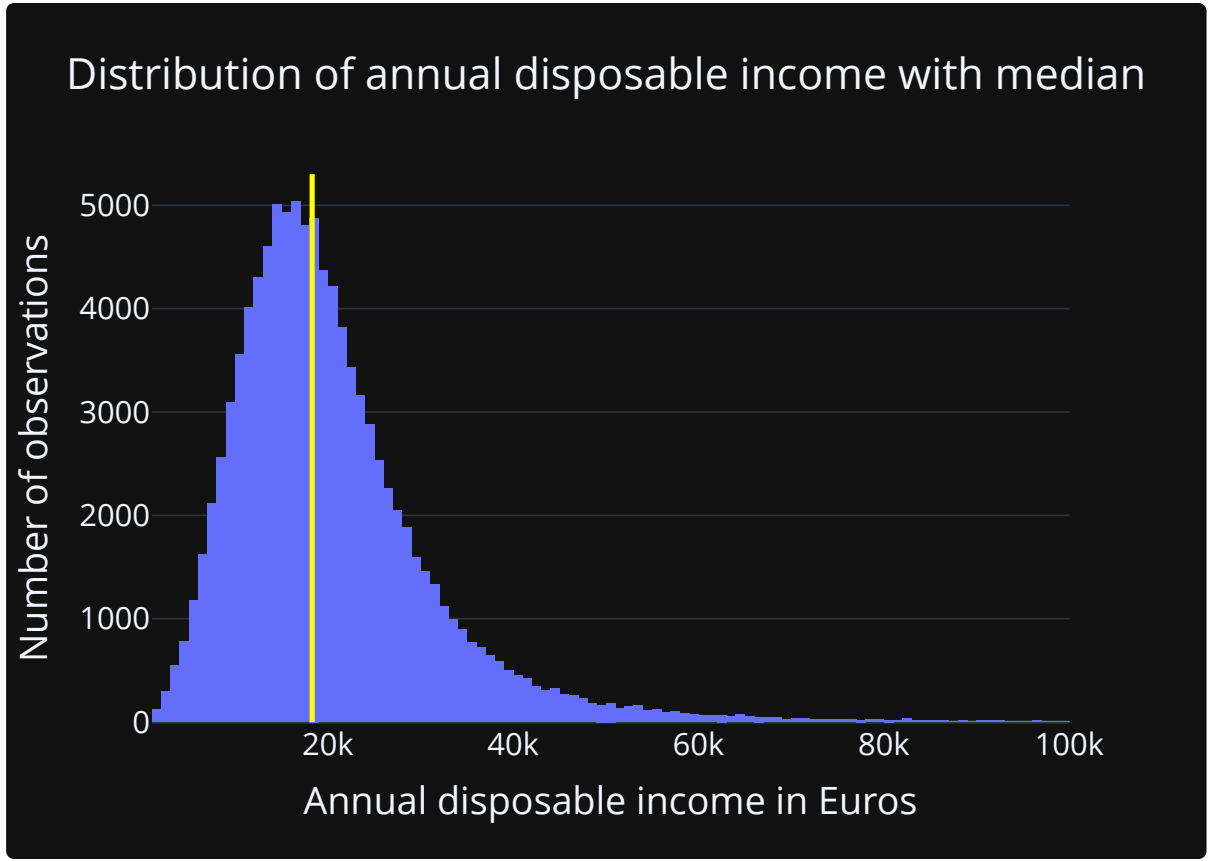
Statistics — Basics & location

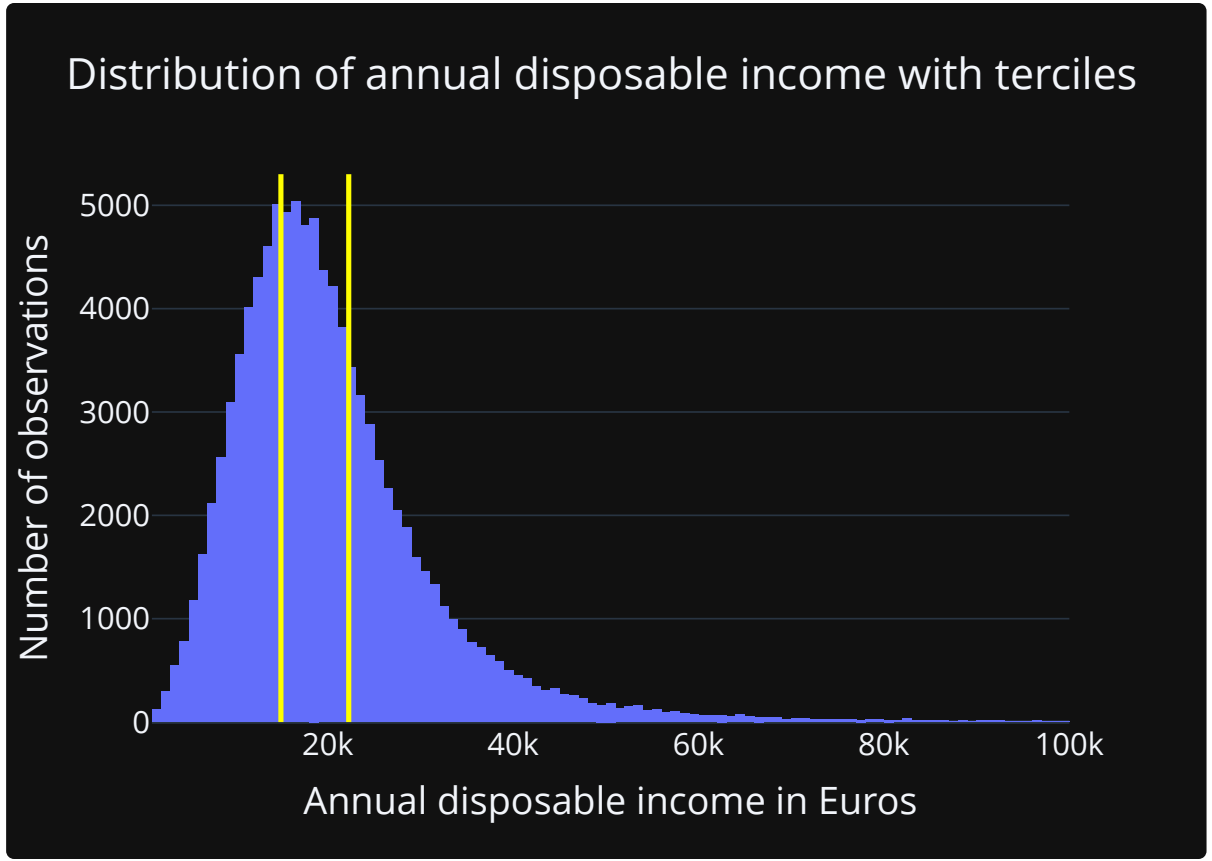
Quantiles

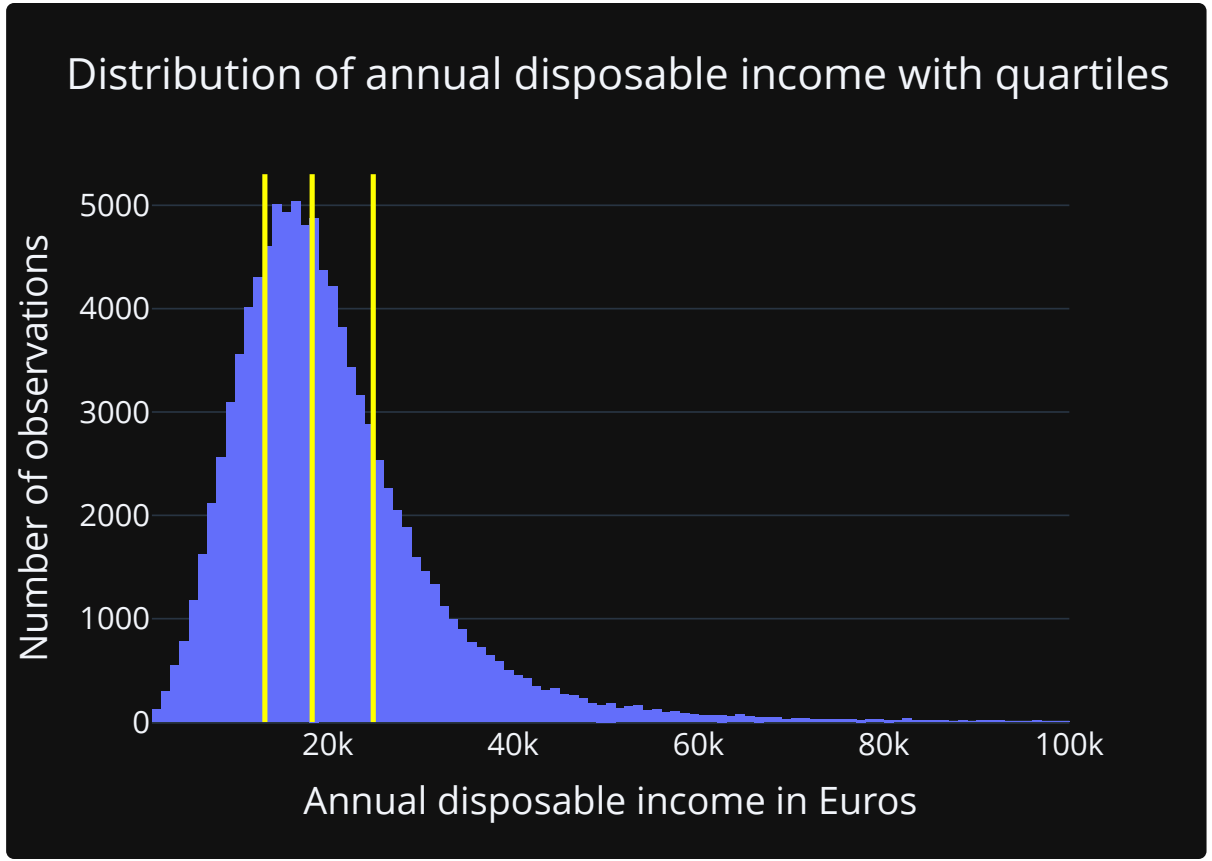
Hans-Martin von Gaudecker and Aapo Stenhammar

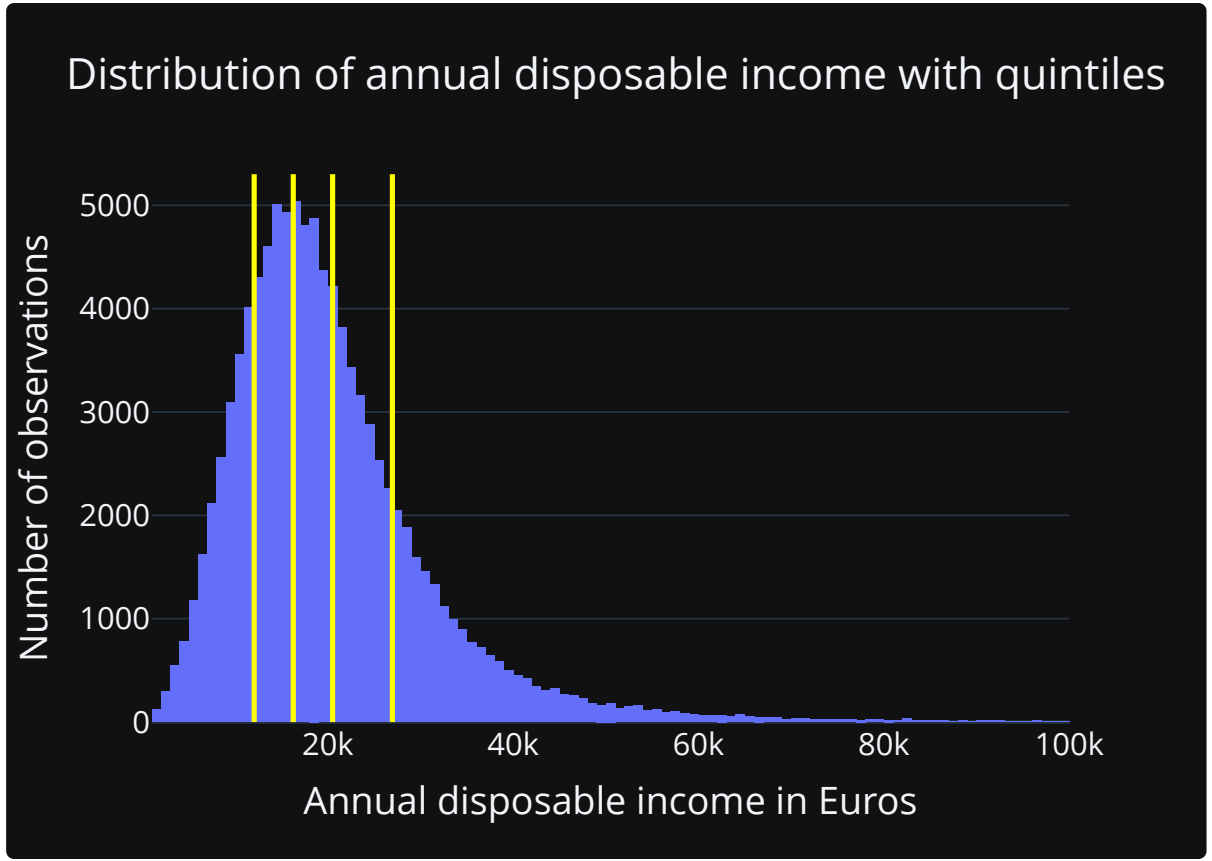
- Median divides the data into two halves
- Terciles divide the data into three equally sized bins
- Quartiles divide the data into four equally sized bins
- Quintiles divide the data into five equally sized bins
- Deciles divide the data into ten equally sized bins
- Percentiles divide the data into one hundred equally sized bins

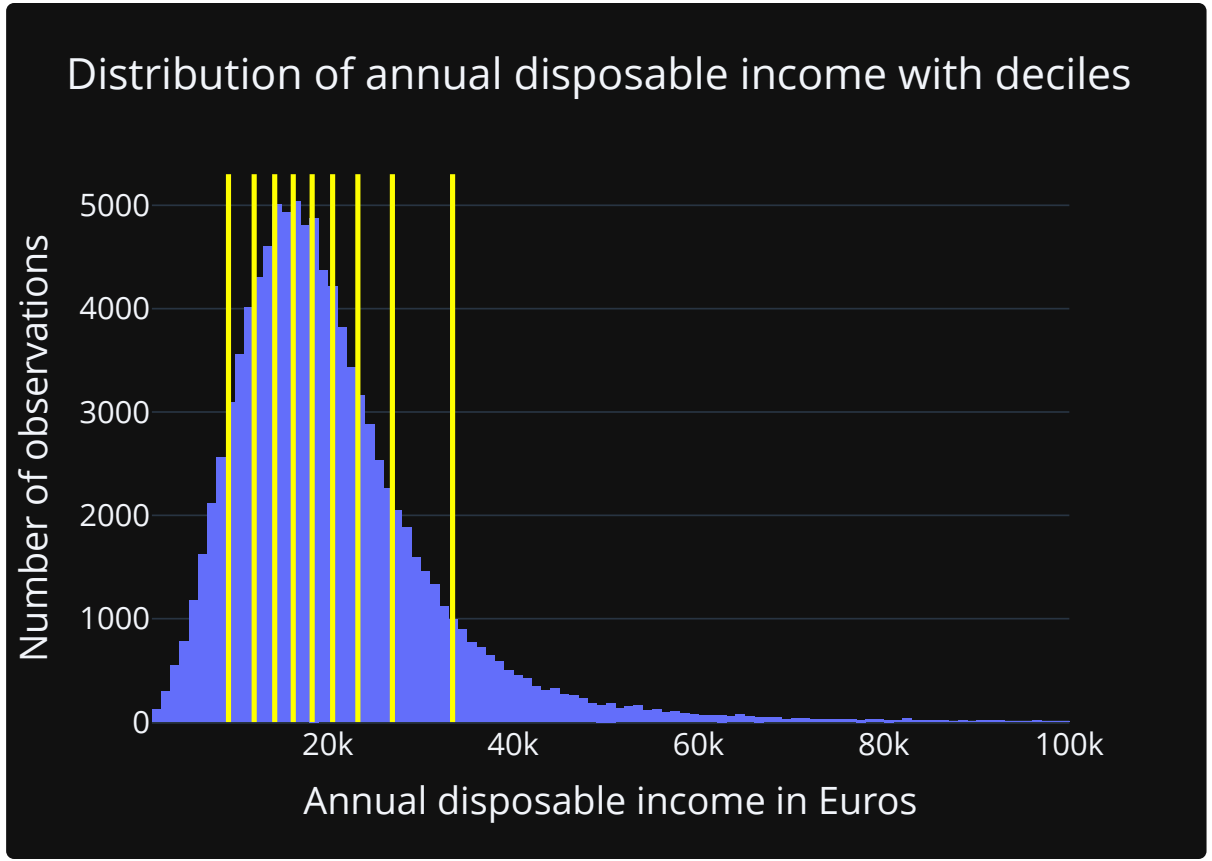

Distribution of annual disposable income
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros

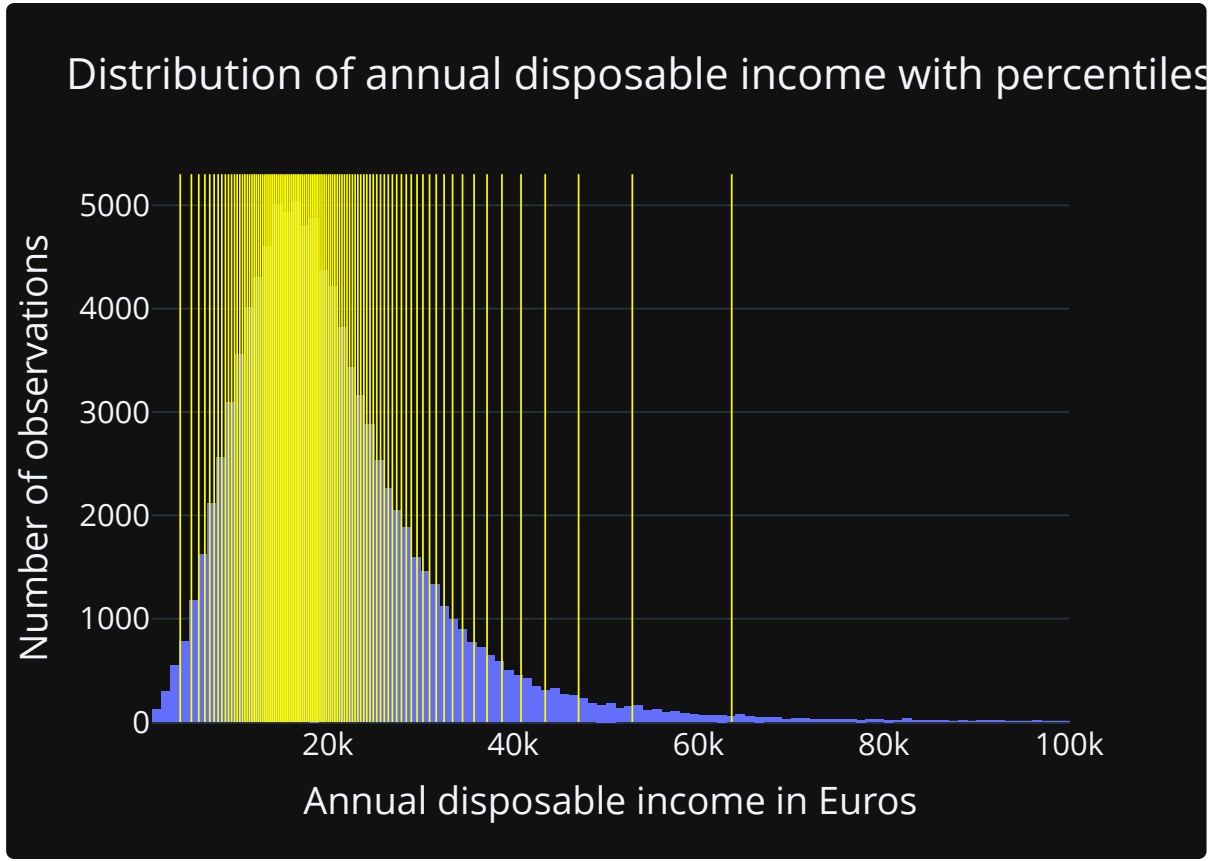

Distribution of annual disposable income with median
Number of observations
5000
4000
3000
2000
1000
0
20k
40k
60k
80k
100k
Annual disposable income in Euros


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

Quantile $\in [0, 1]$

The q-th quantile of a distribution is the value x such that:

1. A fraction q of the data is less than or equal to x.
2. A fraction $1 - q$ is greater than or equal to x.

Examples:

- The median is the 0.5 quantile.
- The 57th percentile is the 0.57 quantile.

Terminology often a bit loose

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

Another example: "Observations in the 90th percentile" would mean observations such that

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

Hence, reference to bins instead of points.