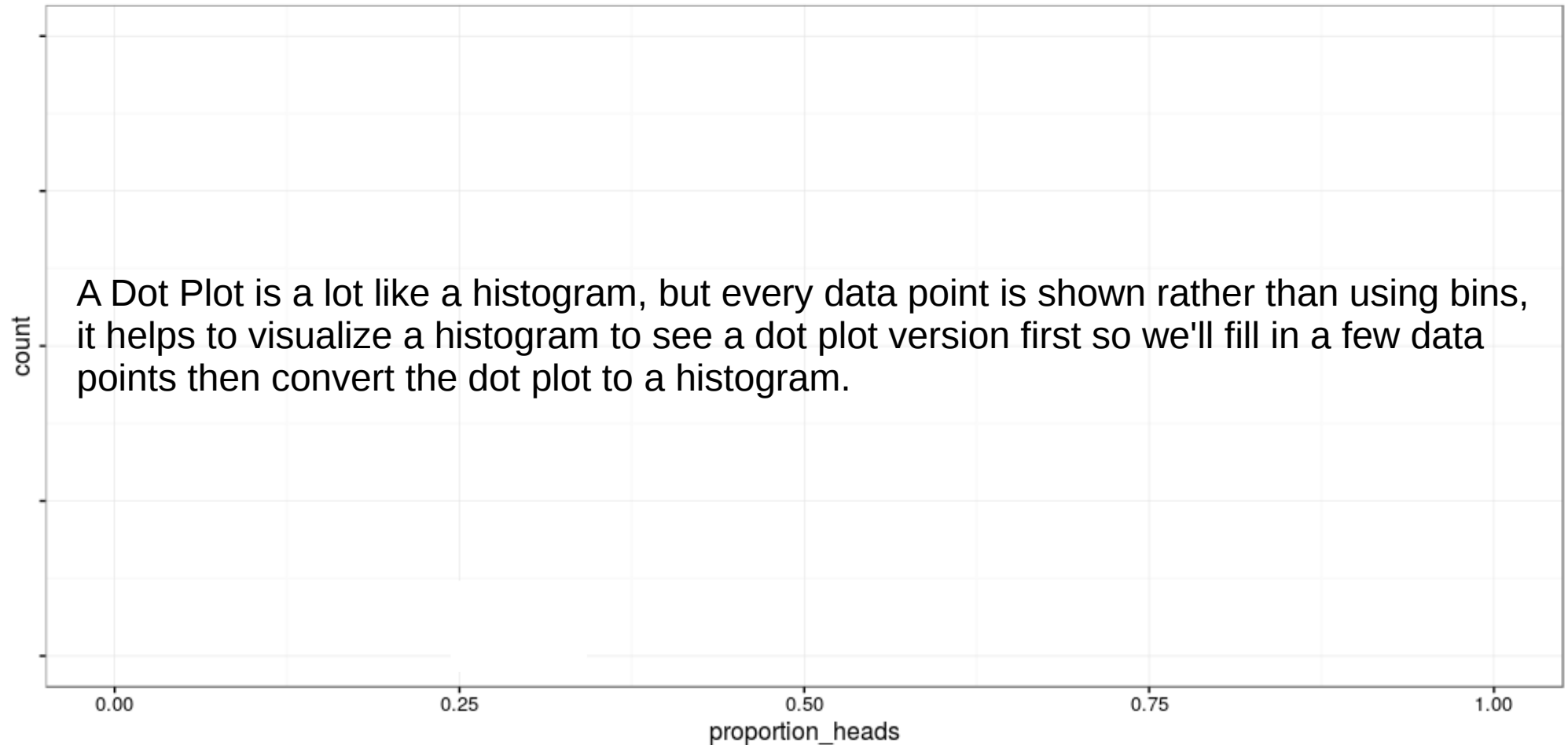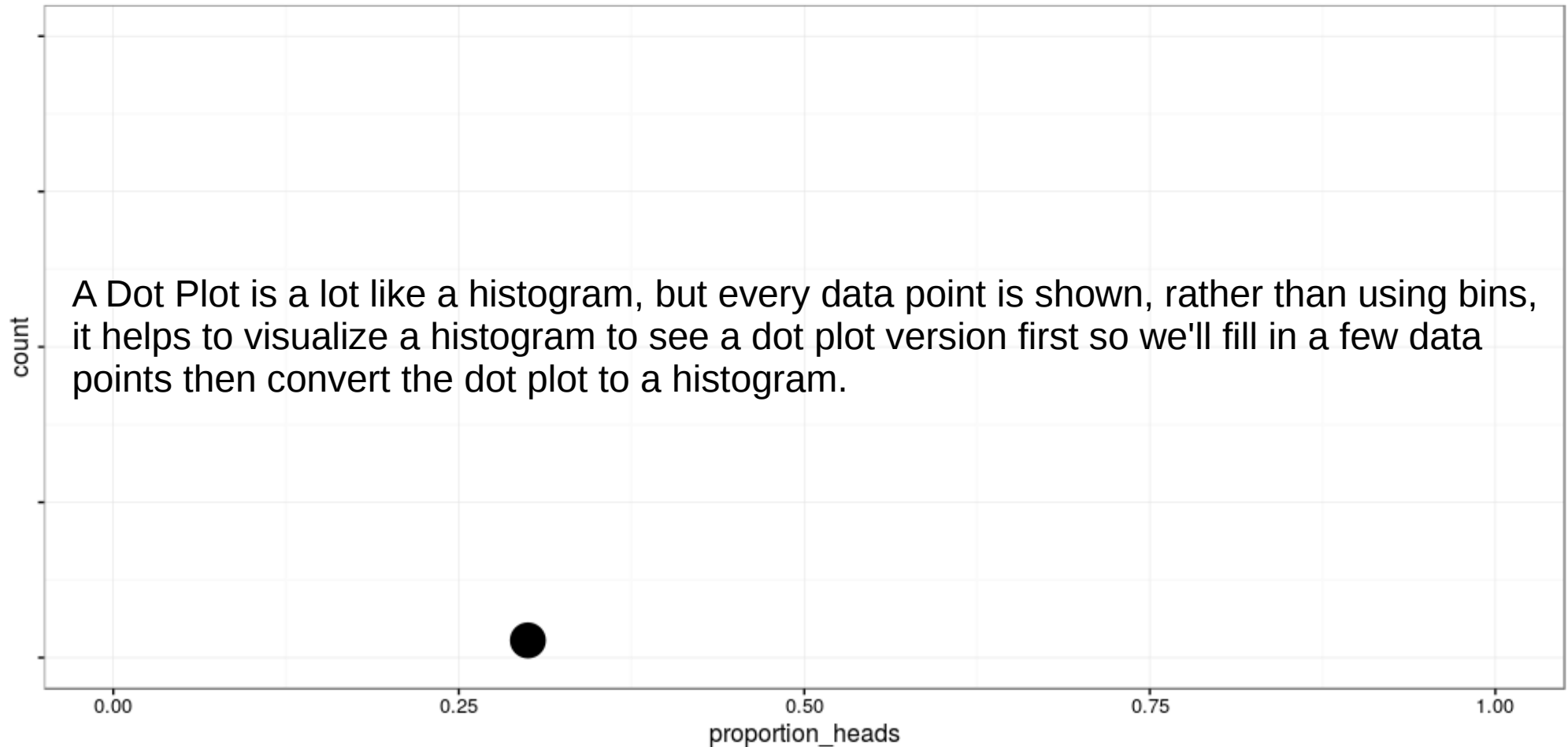**Histograms** – show you the distribution of your data. They have a binned numerical variable on the X axis and count on the Y axis.

Histograms – show you the distribution of your data. They have a binned numerical variable on the X axis and count on the Y axis.

A Dot Plot is a lot like a histogram, but every data point is shown rather than using bins, it helps to visualize a histogram to see a dot plot version first so we'll fill in a few data points then convert the dot plot to a histogram.

count

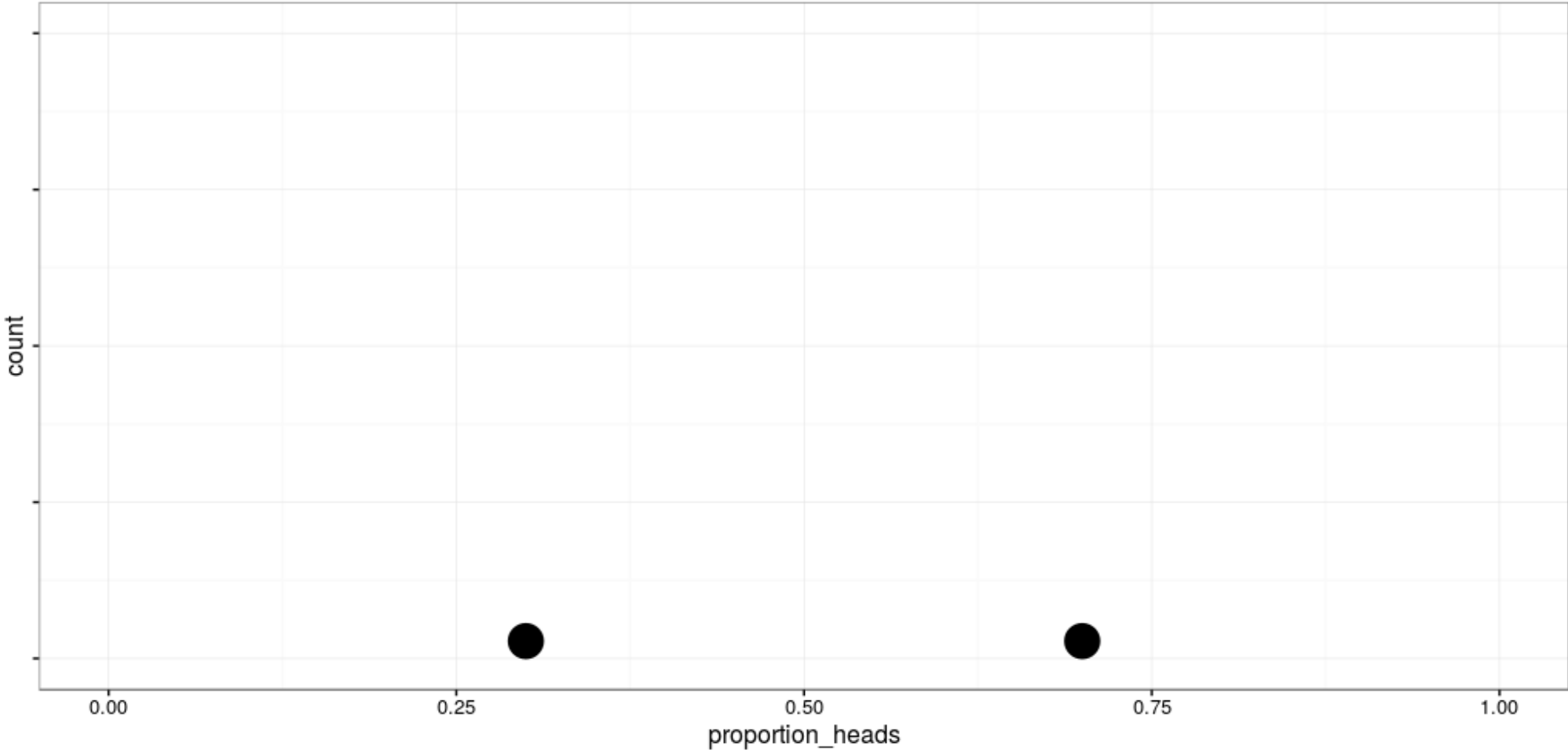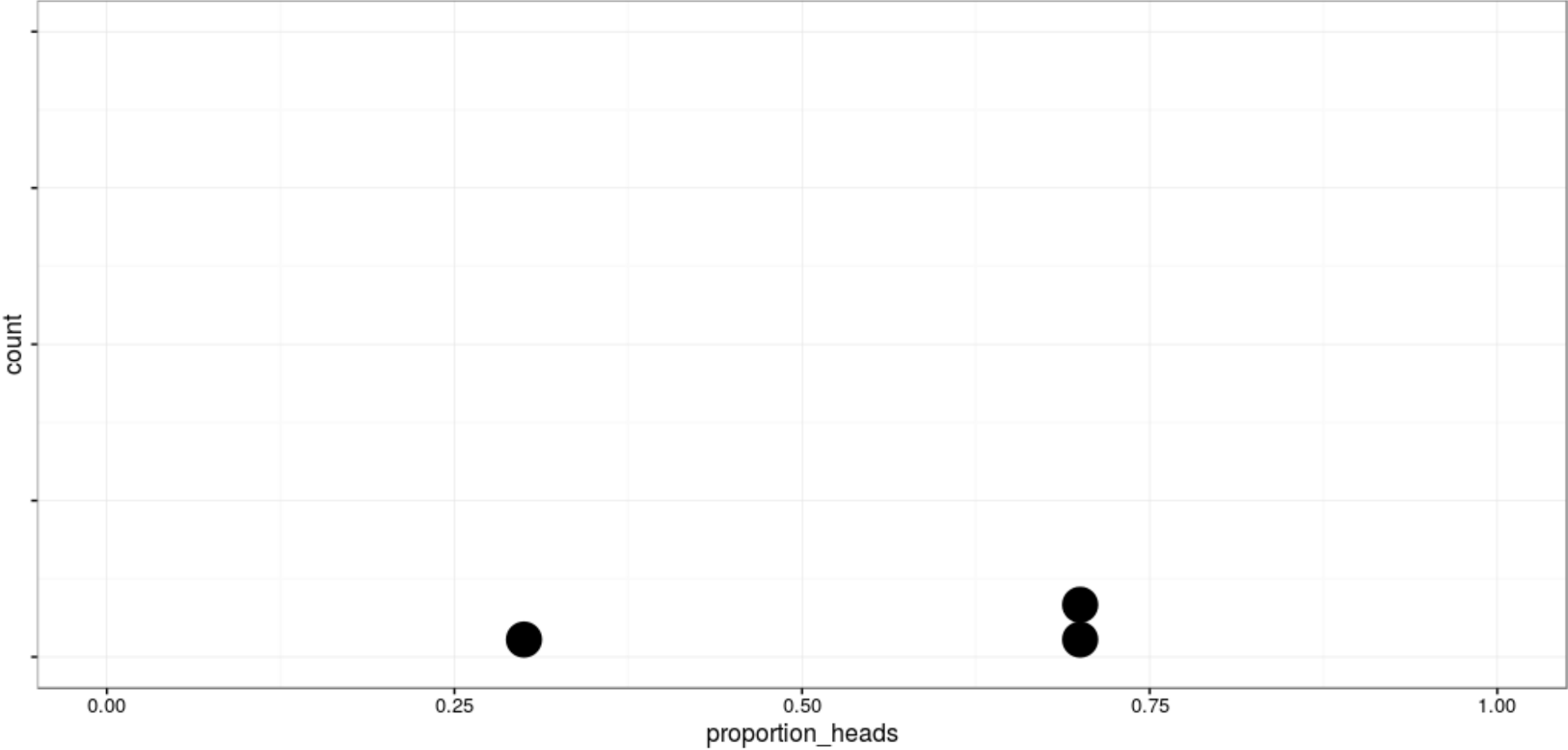0.00   0.25   0.50   0.75   1.00
proportion_heads

Histograms – show you the distribution of your data. They have a binned numerical variable on the X axis and count on the Y axis.
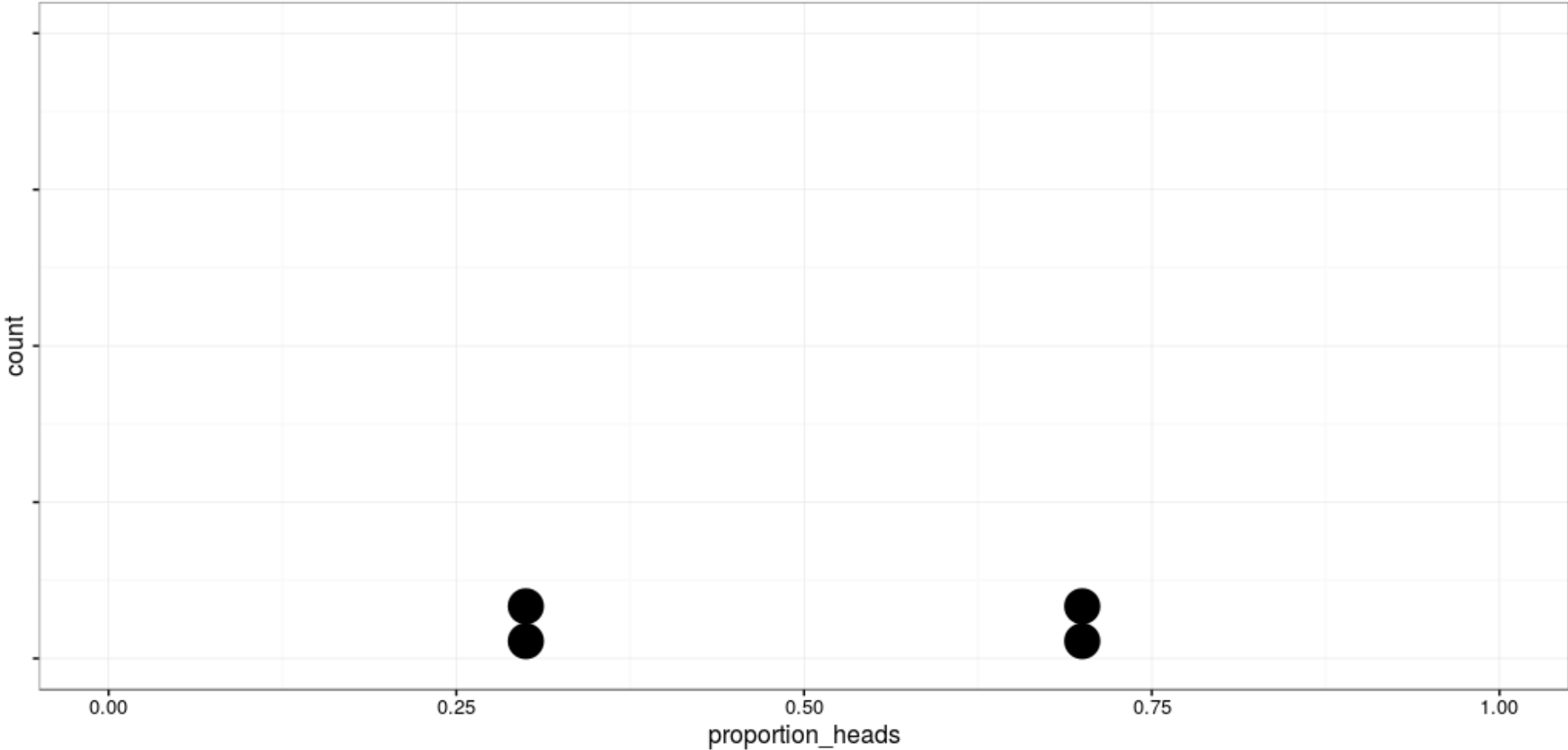
Plot our first data point.

A Dot Plot is a lot like a histogram, but every data point is shown, rather than using bins, it helps to visualize a histogram to see a dot plot version first so we'll fill in a few data points then convert the dot plot to a histogram.
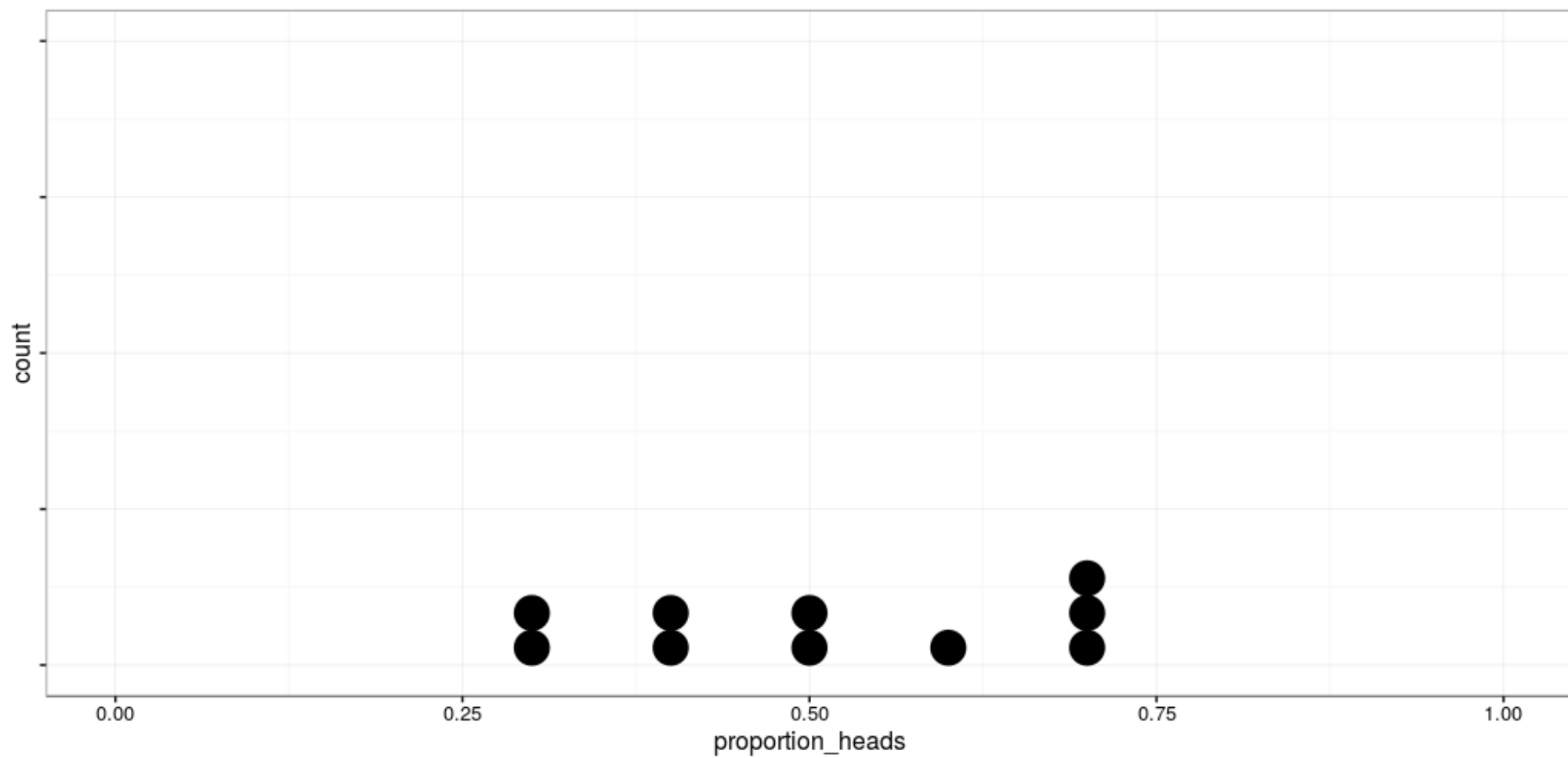
Second data point.
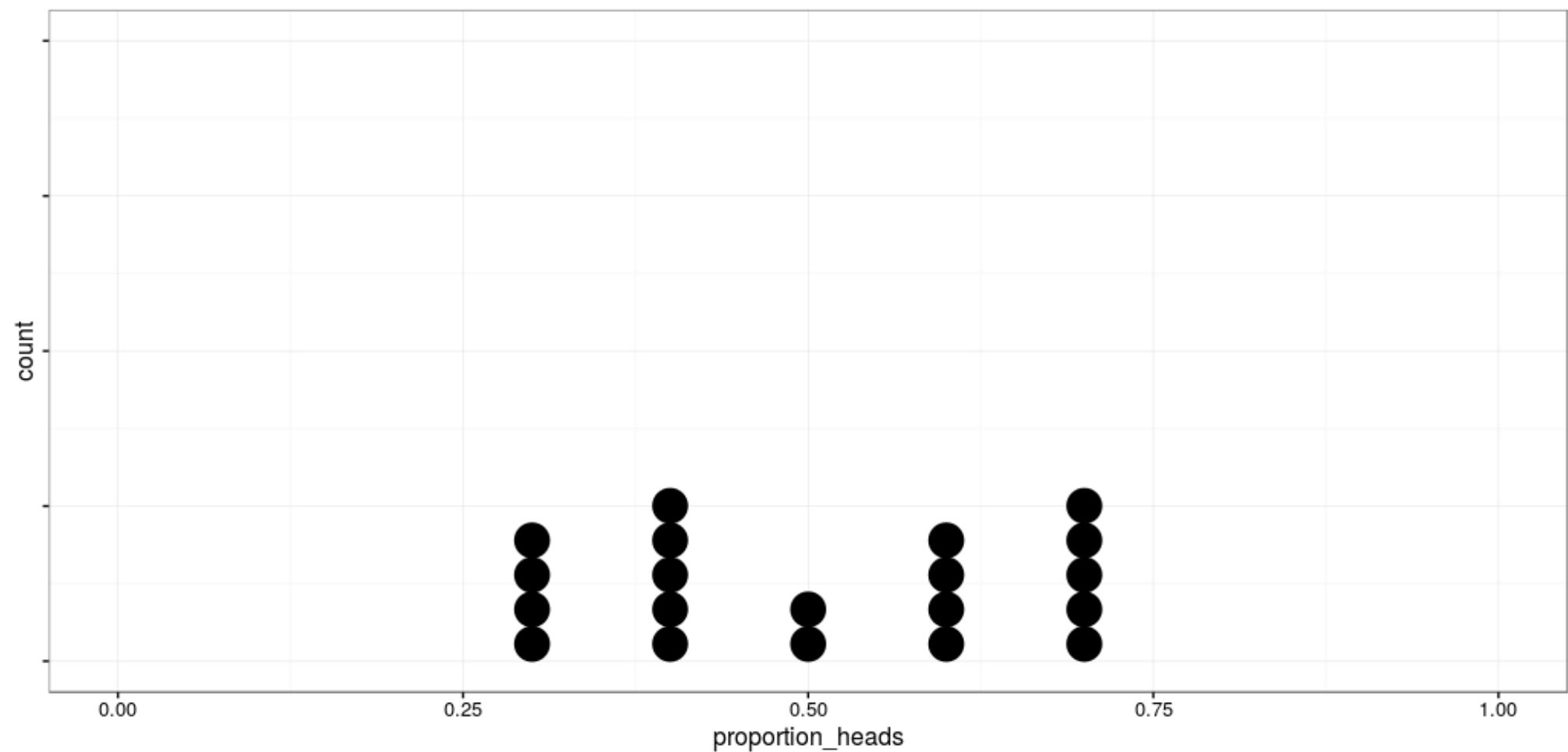
Third data point.
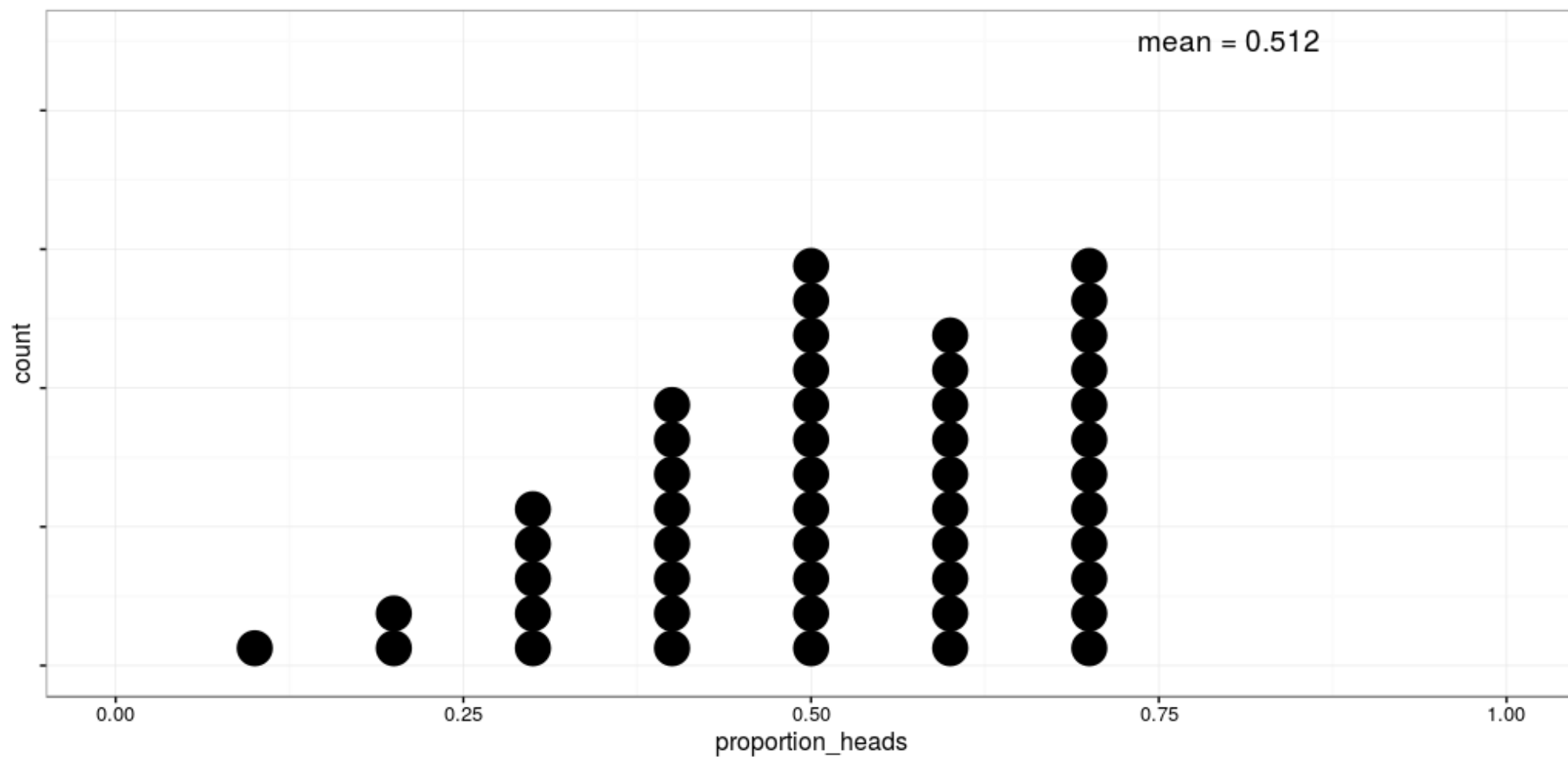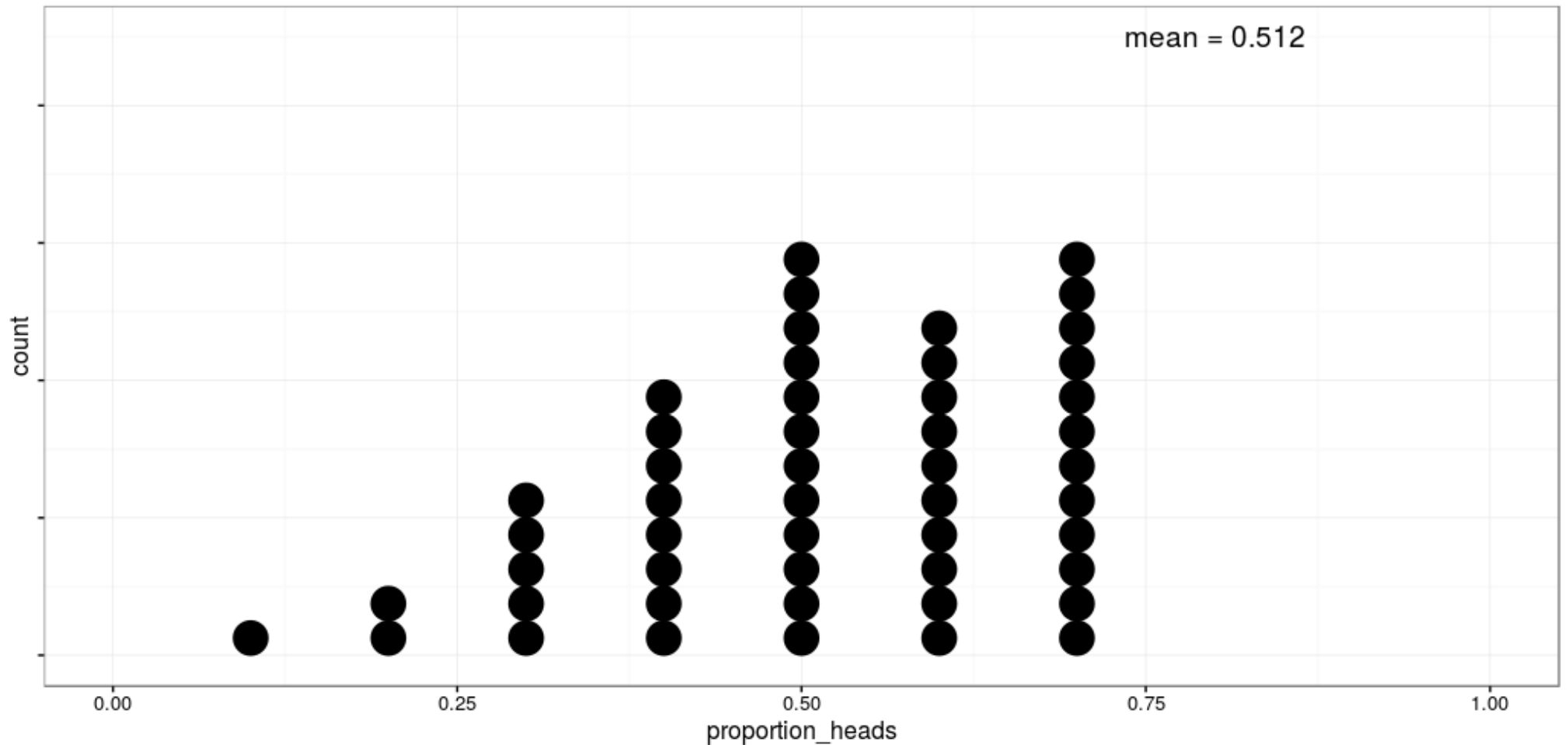
Fourth data point.

Ten data points.

Twenty data points.

All fifty data points on a dot plot.

All fifty data points on a dot plot.

Now let's convert this to a histogram by binning the data. The nice thing is, that since this data was generated using 10 coin flips for each data point, there are a finite number of possibilities and the bins are pretty obvious.
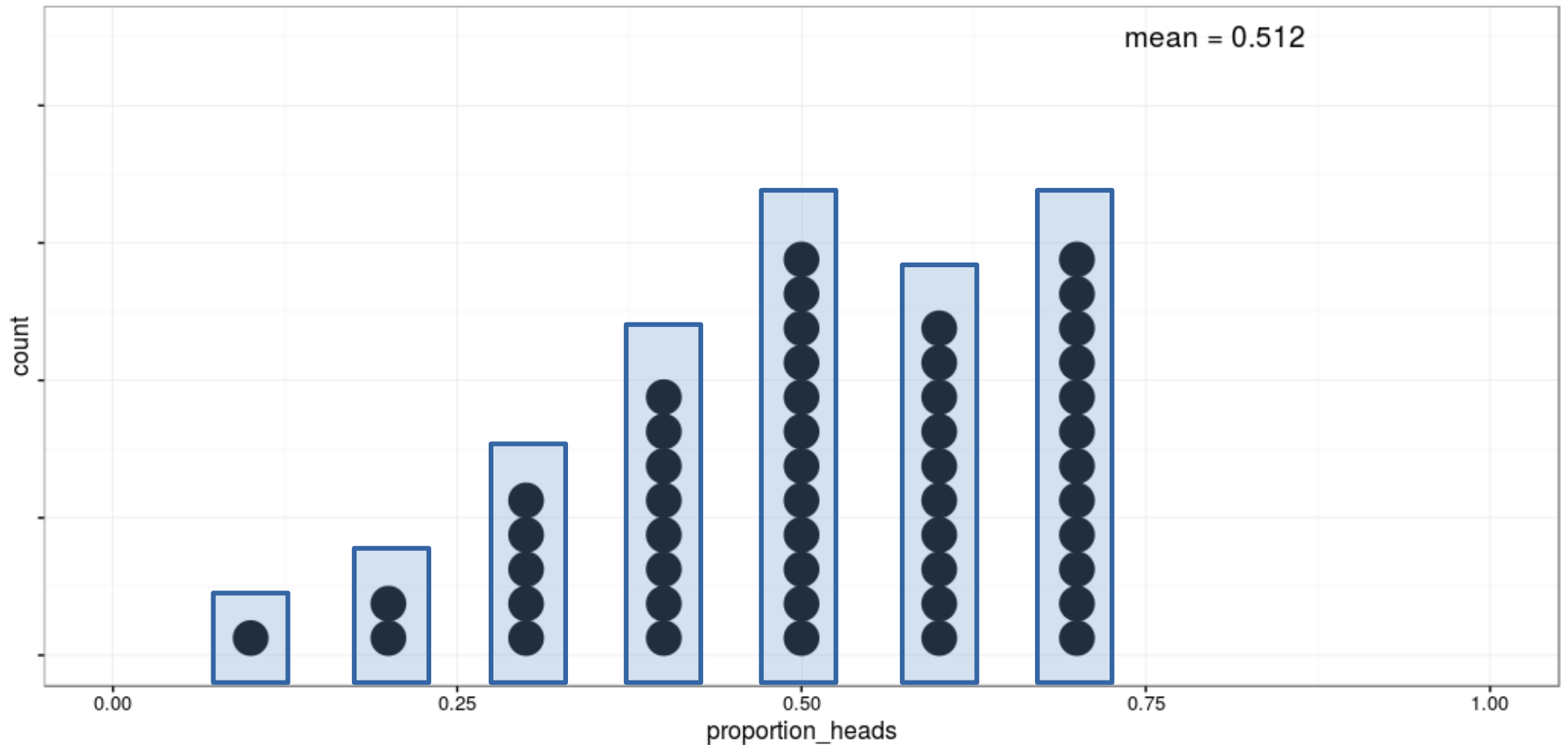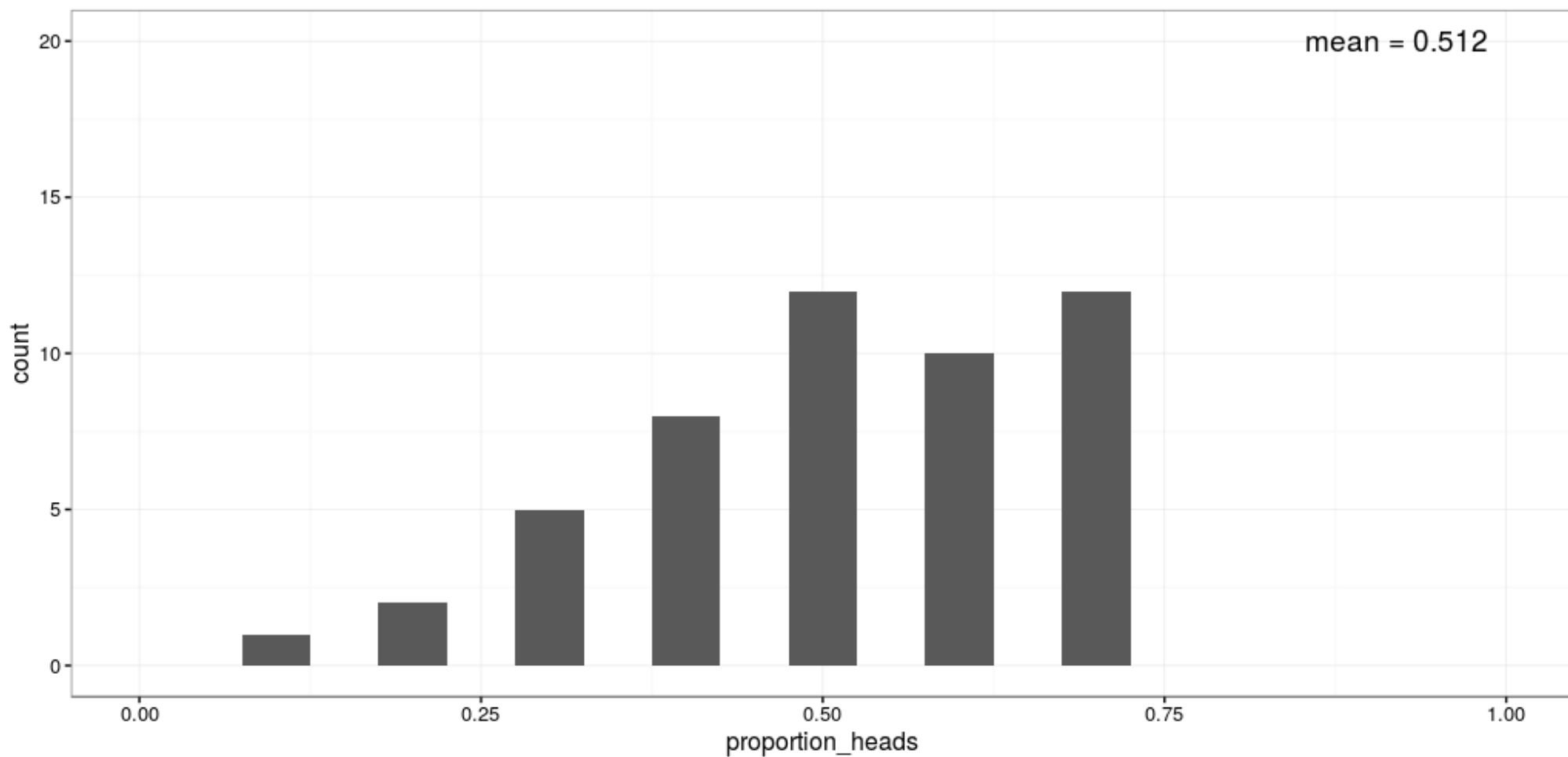
All fifty data points on a dot plot.

Now let's convert this to a histogram by binning the data. The nice thing is, that since this data was generated using 10 coin flips for each data point, there are a finite number of possibilities and the bins are pretty obvious.
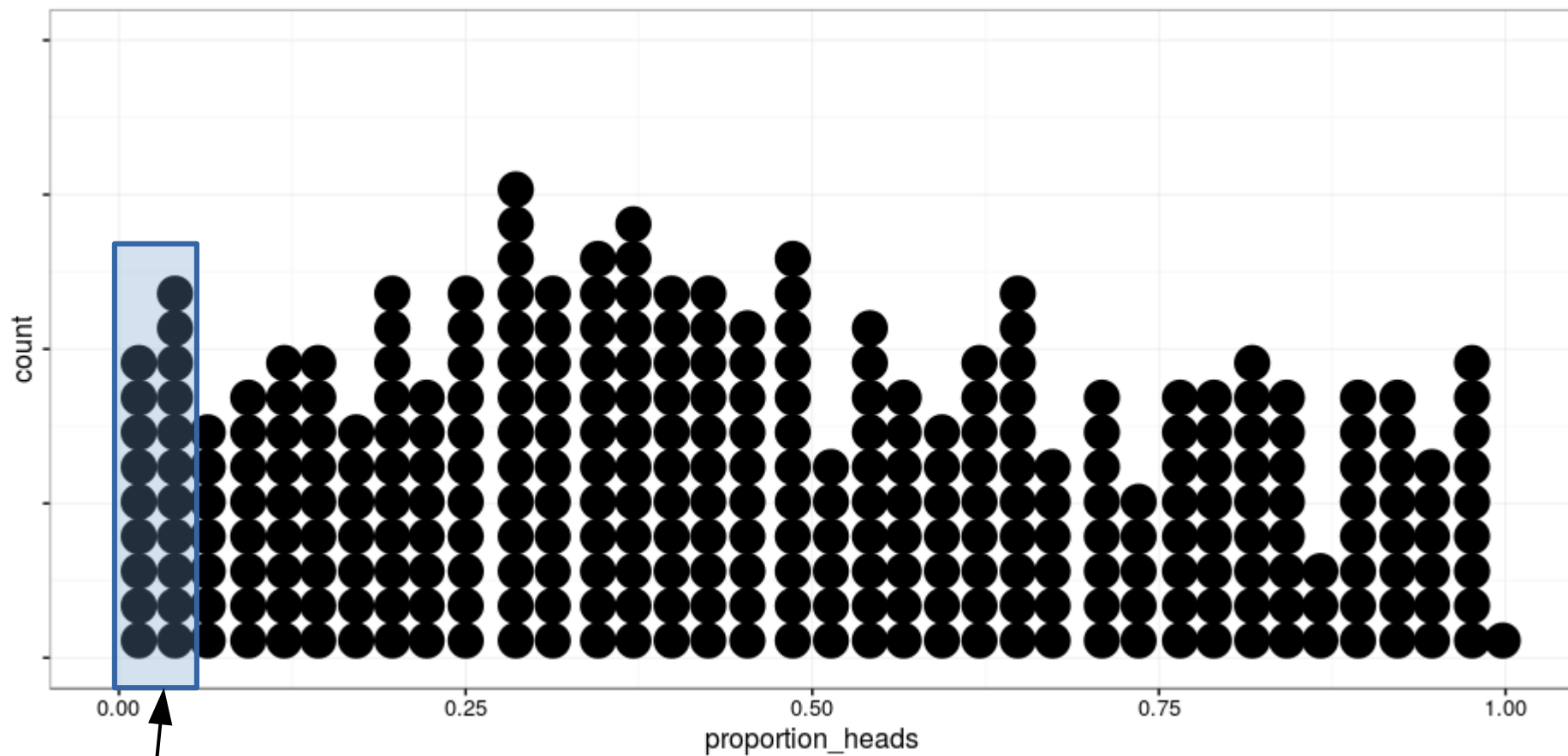
Final Histogram of the same data that is in the dot plot of 50 data points.

Dot plots start to get really busy, and binning to make a histogram is really better.
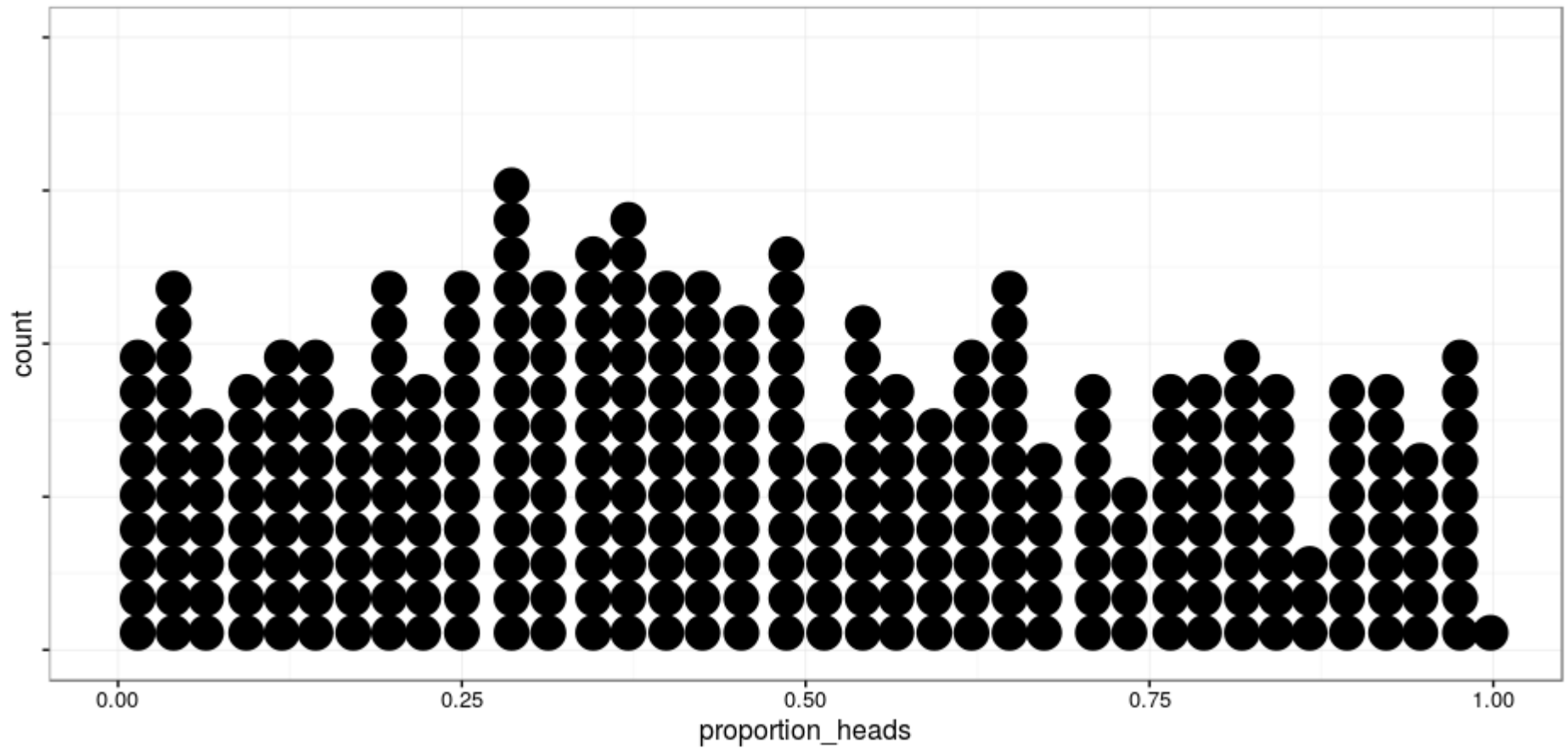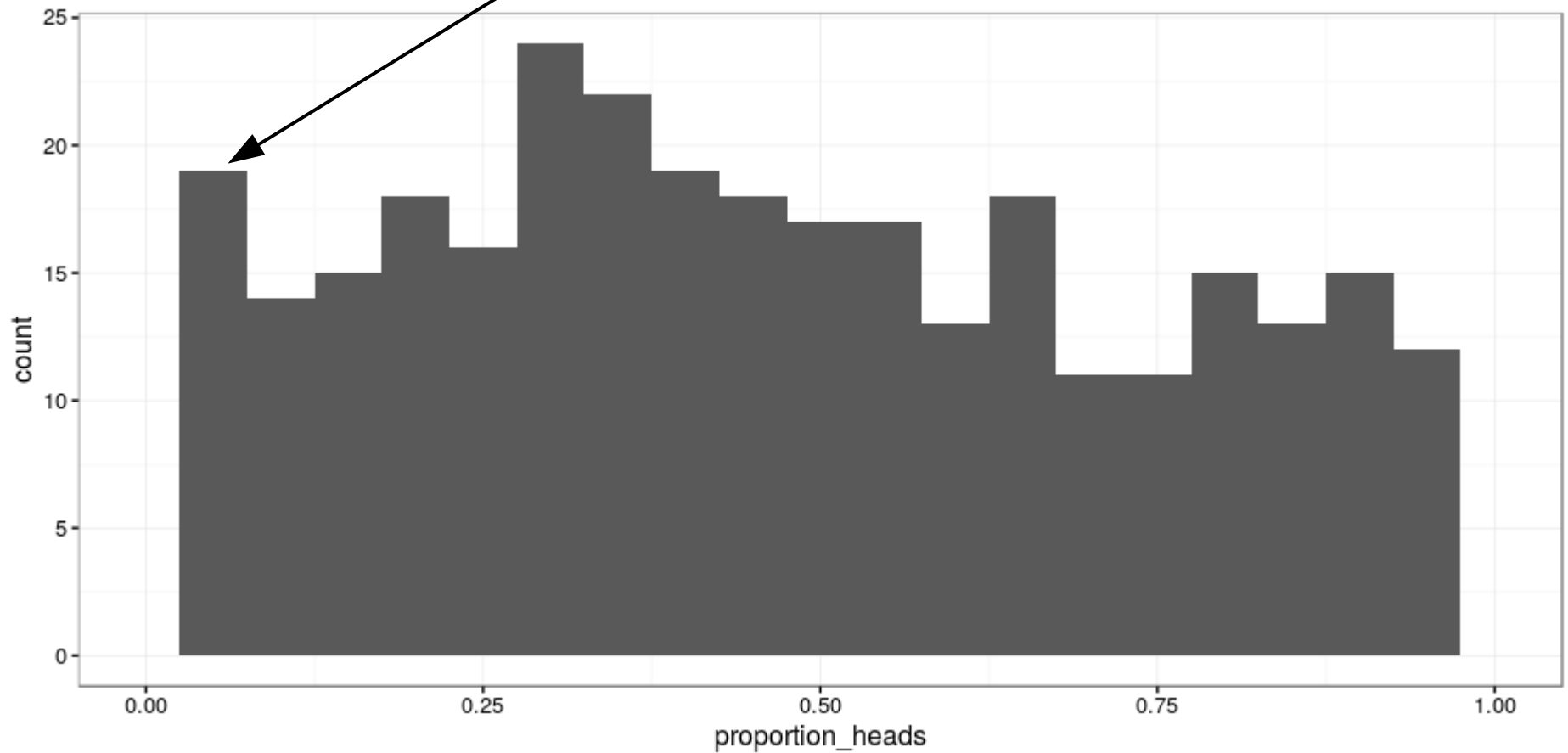
Dot plots start to get really busy, and binning to make a histogram is really better.
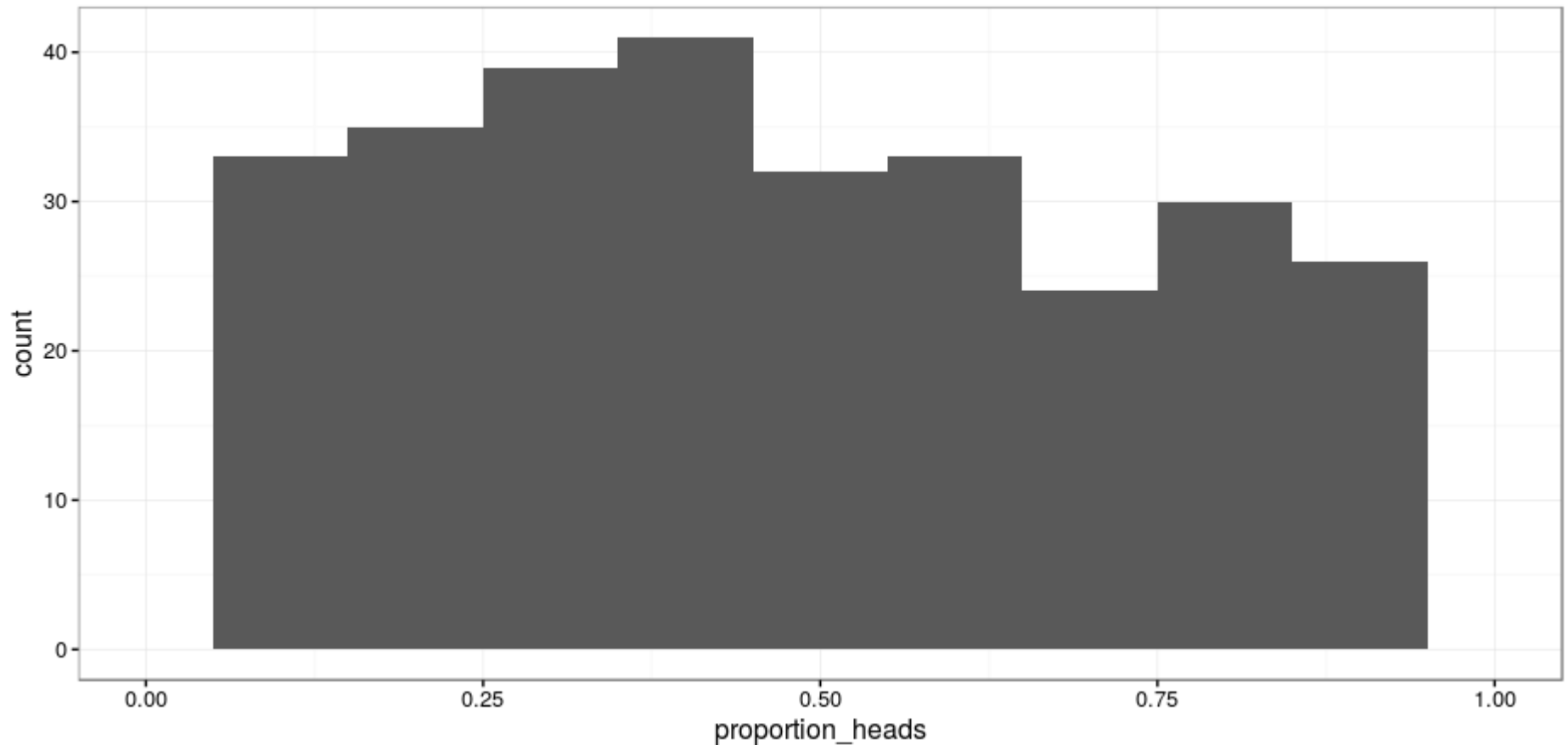

example bin

See? This histogram has a bin width of 0.05 and combines about 2 columns of dots from the previous dot plot into a bin.

This histogram has a bin width of 0.1 and combines more columns of dots from the previous dot plot into a bin. Choosing a bin width isn't always easy, but you want to accurately represent the distribution of your data. (these are a bit too big for me)

Stacked histogram – it's like 3 histograms for the price of one! And lets you compare the distributions of a numerical variable (weight) among the categories (genotype)
All 3 of the categories share the same X axis, and the left hand Y axis (count) repeats on the same scale for each category

weight distribution by genotype

This is the numerical variable (weight)
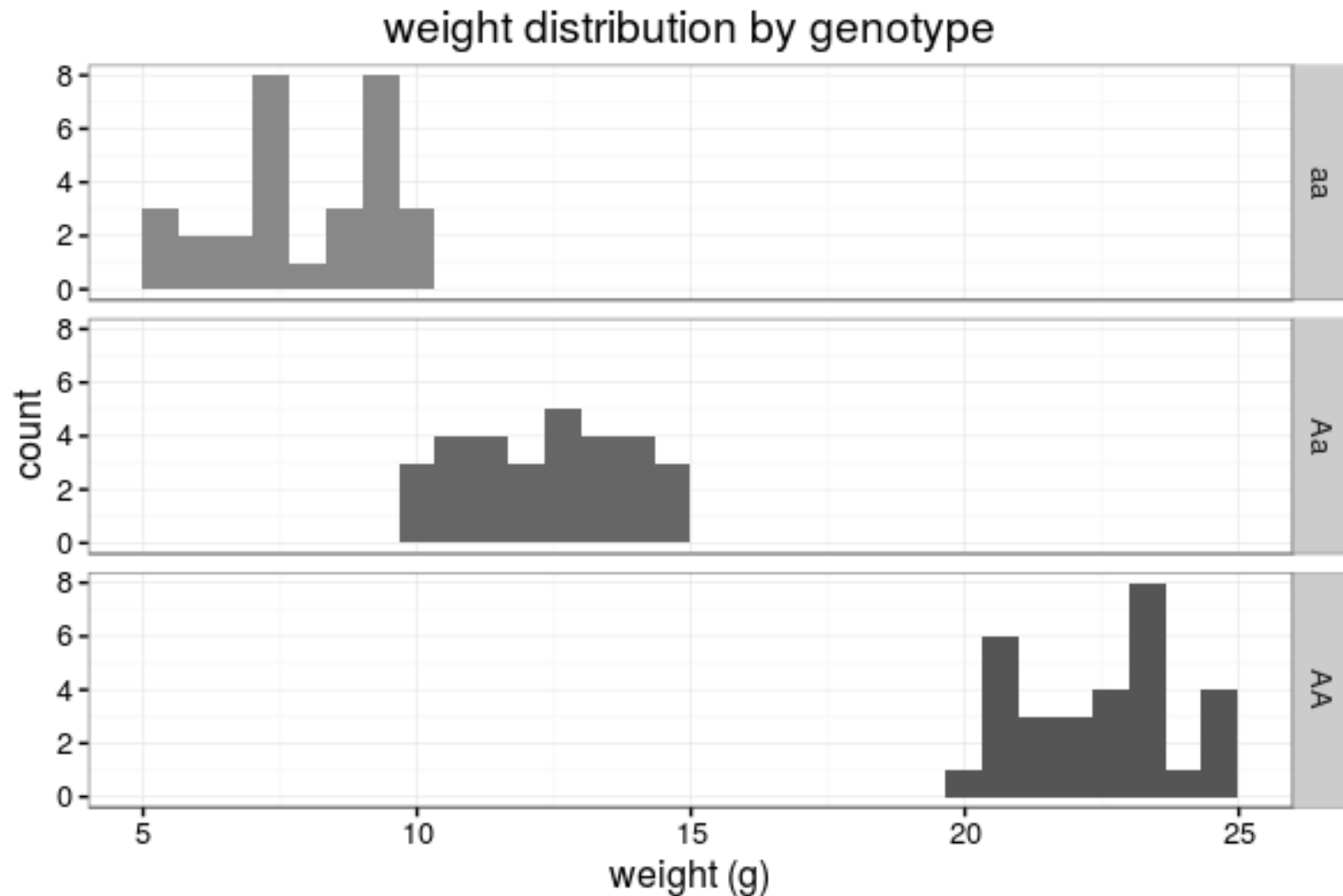
This is the categorical variable (genotype)

Box Plot – if you want to provide a little more summary information about the numerical values for each category, a side-by-side box plot is nice.
It also lets you compare the distributions and medians of a numerical variable (weight) among the categories (genotype). All 3 of the categories share the same Y axis (weight) with each category having its own box.
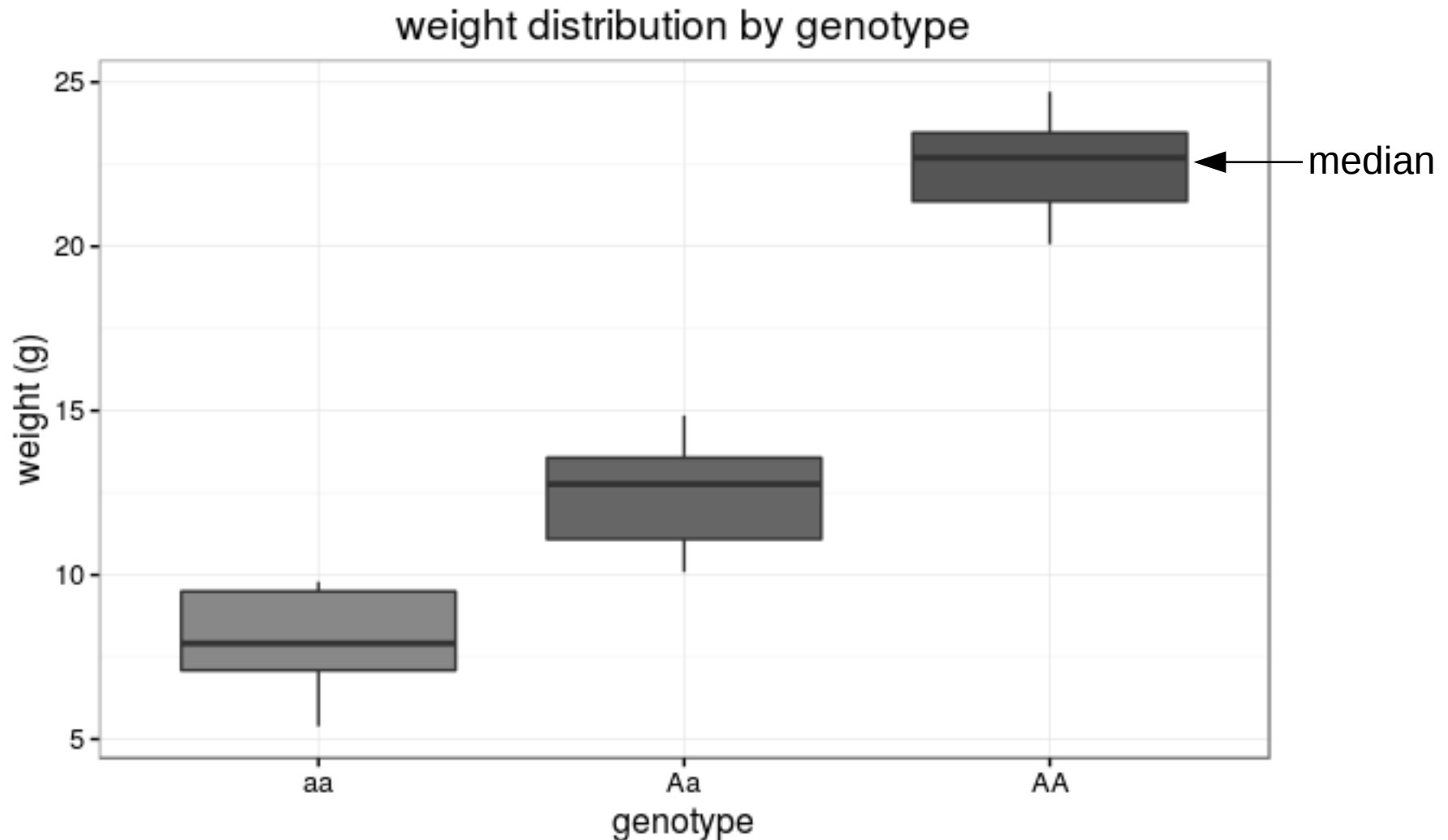

weight distribution by genotype

Fig 4.12 Description of a Boxplot

Outlier   --------➤ o

←-------- Upper Extreme

←-------- Upper Hinge (Quartile)

←-------- Median

←-------- Lower Hinge (Quartile)

←-------- Lower Extreme