**. Box plots**

Individual histograms are great, but there is a problem if you want to draw lots of them.

**2. You can't just color in histograms**

Let's revisit the kings and queens dataset. Suppose we want to see the distribution of ages for each royal house. A naive solution might be to draw the same histogram, but using different colors for each house. Sadly, this is a horrible muddled mess.

**3. Draw each histogram in its own panel**

In many cases, the only sensible way to draw lots of histograms is to draw them in their own panel. Here you can see thirteen histograms, one for each house. This approach still has problems.

**4. Draw each histogram in its own panel**

It's quite easy to compare distributions for panels that are in the same column. You can see that monarchs from the Wessex family were typically much younger when they began ruling than those from the Windsor family, since you can look down the column and see that the Wessex distribution is to the left of Windsor's. By contrast, it's harder to compare distributions between panels that are in different columns.

**5. Draw each histogram in its own panel**

To compare the ages of monarchs in the rival York and Lancaster houses, you have to do a lot of looking back and forth and staring at numbers on the x-axis, which isn't ideal.

**6. 1-col**

You could align all the panels in a single column, but that often means running out of space. Here, the text on the plot is almost unreadable. Fortunately, box plots can solve our problems.

**7. When should you use a box plot?**

Box plots split a continuous variable - like age - by a categorical variable - like royal house - and allow us to compare the resulting distributions in a space-efficient way.

**8. Histogram vs. box plot**

Here's a comparison of the histogram you saw before with a box plot.

**9. Histogram vs. box plot: mid-line**

The line in the middle shows the median of the distribution. That is, half the monarchs started ruling before this age, and half after this age.

**10. Histograms vs. box plot: the box**

The box in the box plot extends from the lower quartile to the upper quartile. The lower quartile is the point where one quarter of the values are below it. That is, one quarter of the monarchs started ruling before this age, and three quarters after it. Likewise, the upper quartile is the age where three quarters of the monarchs started ruling below this age. The difference between the upper quartile and the lower quartile is called the inter-quartile range.

**11. Histograms vs. box plots: the whiskers**

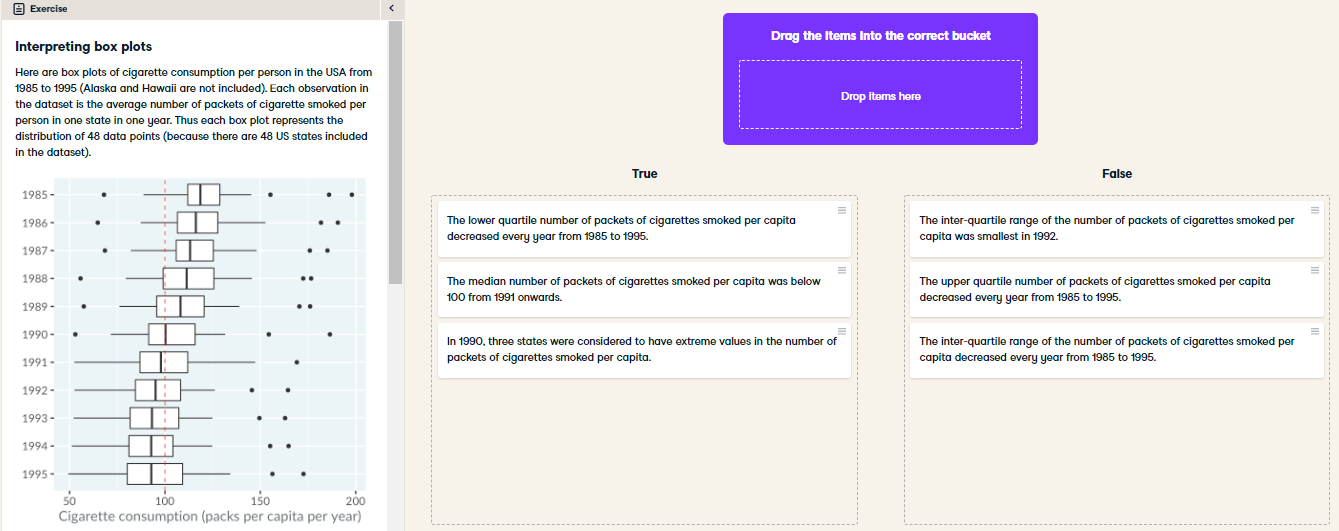
The horizontal lines, known as "whiskers", have a more complicated definition. Each bar extends to one and a half times the interquartile range, but then they are limited to reaching actual data points. The technical definition is shown in the slide, but in practice, you can think of the whiskers as extending far enough that anything outside of them is an extreme value.

**12. Monarchs by house**

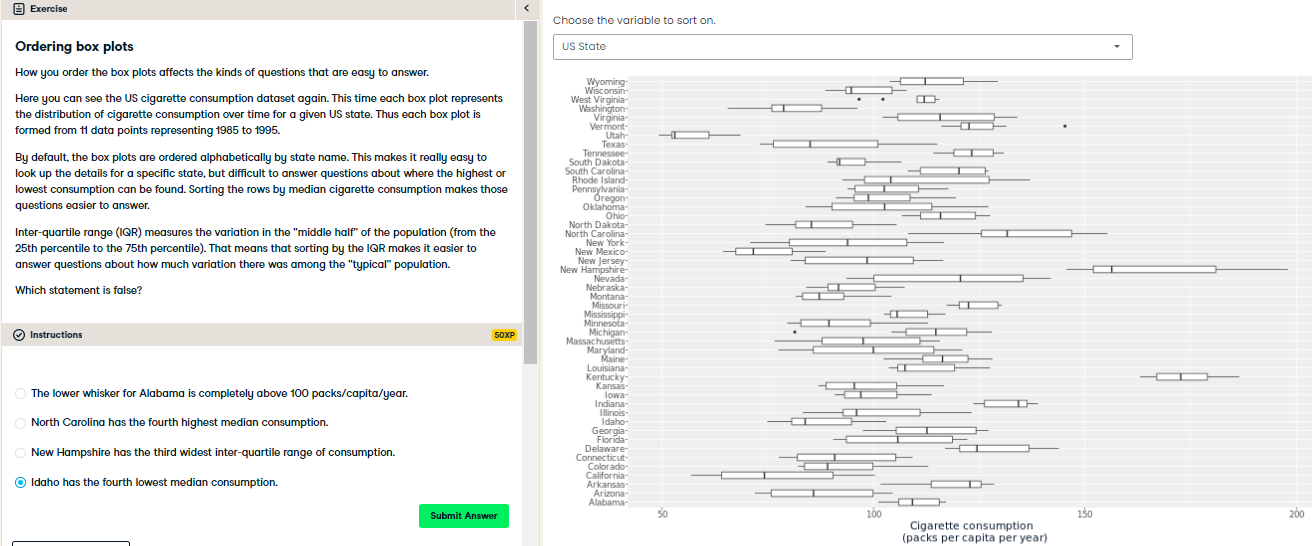
As mentioned before, the power of box plots is that you can compare many distributions at once. Here, the royal houses are ordered from oldest at the top to newest at the bottom. A trend is visible: since the Plantagenets in the fourteenth century, the boxes gradually move right showing that the ages when new monarchs ascend to the throne have been increasing. Godwin and Blois appear as a single line because there was only one king from each house. The Anjou house only had three kings, and forms a box with one whisker, not two. Notice that the box plots for the houses of Denmark and Windsor show some points. These are extreme values, that is, values that are outside the range of the whiskers. Windsor's left-most outlier is Elizabeth the second who ascended at age 26.

**13. Let's practice!**

Let's draw some box plots!



Beautiful box plots! Although each box plot contains less information than a histogram, because they are so compact to draw, it is easy to compare many distributions at once.



Awesome ordering! There is no single “best” way to order the box plots. Different orderings make it easier to answer different questions.