

# An Introduction to Boxplots

Boxplots are a fairly popular visualisation type, but one needs to know how to read them in order to extract useful conclusions. This introductory explainer contains an example scenario, a vis construction, and an answering questions section to get you started.

## Introduction

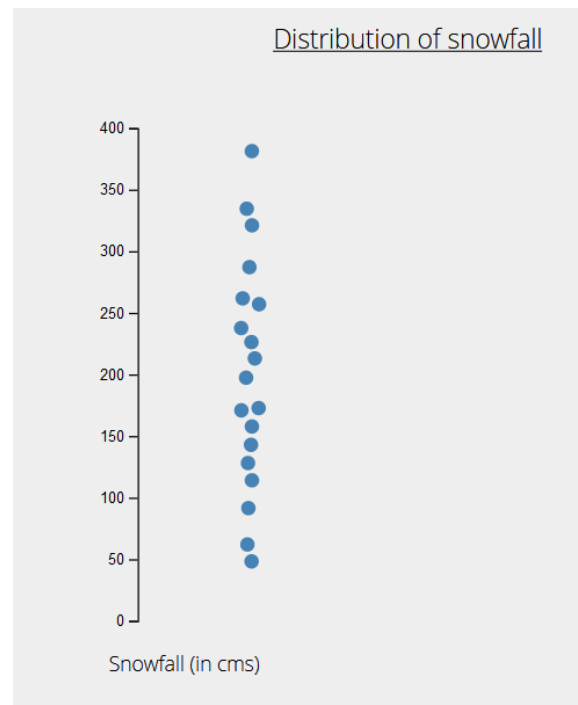
Let's start with an example scenario where boxplots might come in handy...



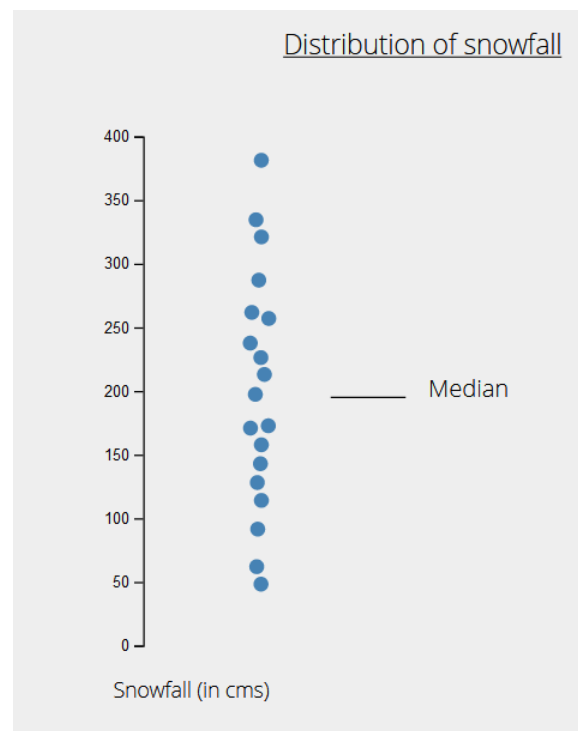
“...why not use a **boxplot**?”

## Construction

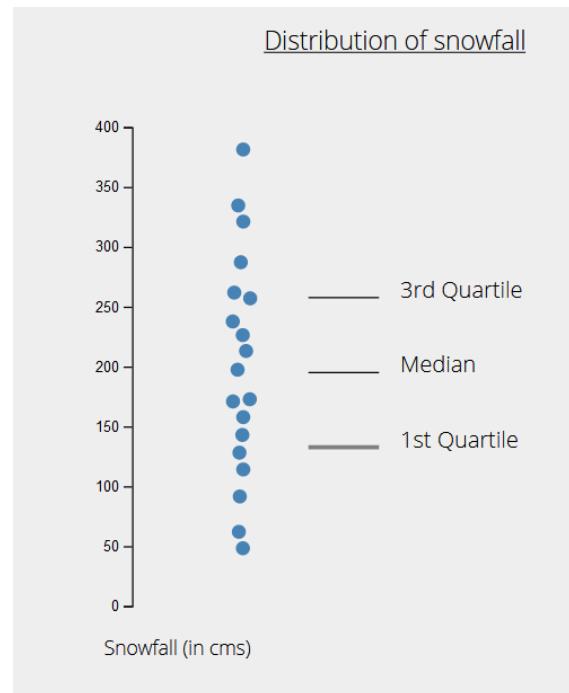
As a first step, let's learn how to construct a boxplot from snowfall data from one mountain. To start, we can **plot the snowfall data for all years**.



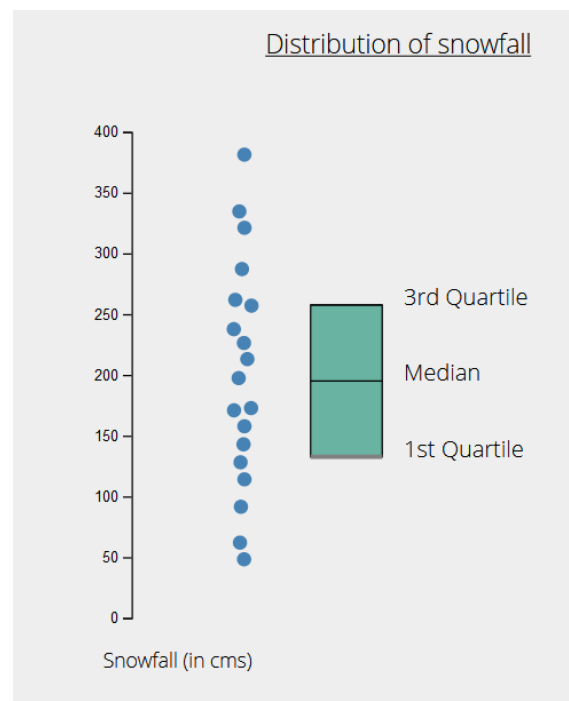
By **splitting the data into two groups with equal number of data points**, we obtain the middle value, called the **median**.



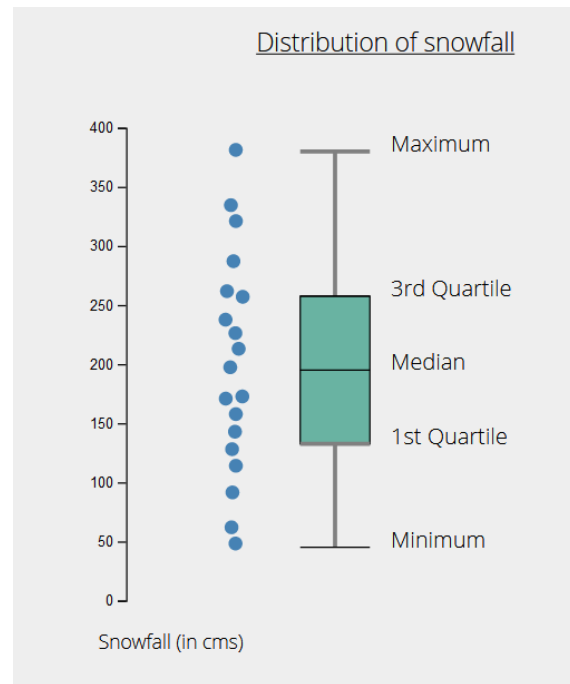
The median values obtained by splitting the lower and higher halves into equal groups themselves are called the **first quartile** and **third quartile**. The quartiles split the data into... quarters.



We link these three values with a **box**, which is called the **interquartile range**. It contains half of the data points.



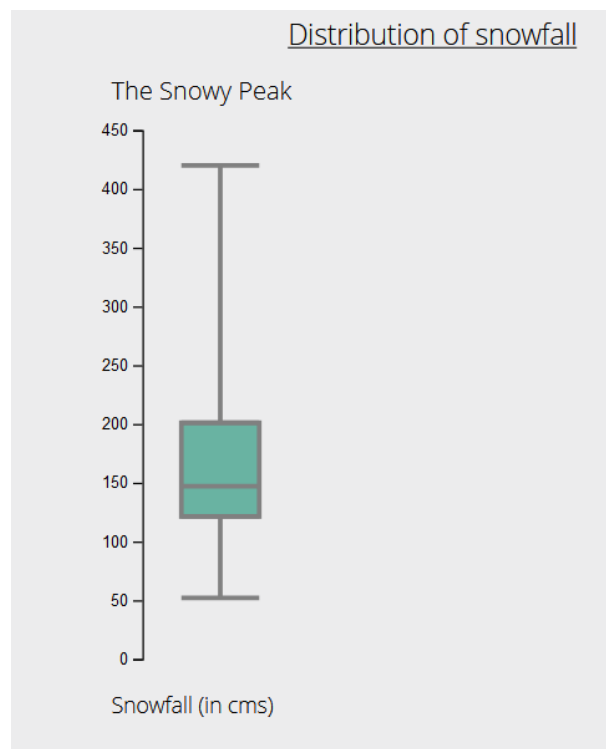
Finally, we plot the **minimum** and **maximum**, which make the "whiskers" of our "box-and-whisker" plot, or boxplot!



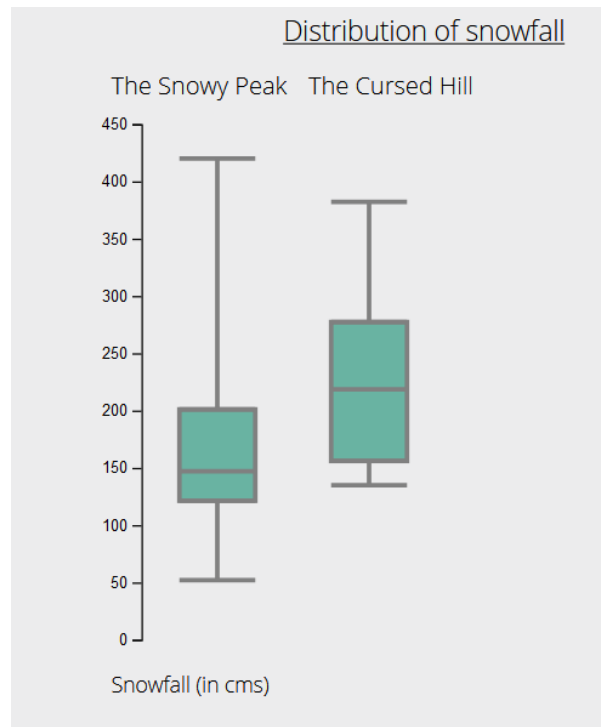
## Answering questions

Great! We now understand how boxplots work. But how do we use boxplots to make a decision? First, we need to plot data from another mountain. Then we can compare.

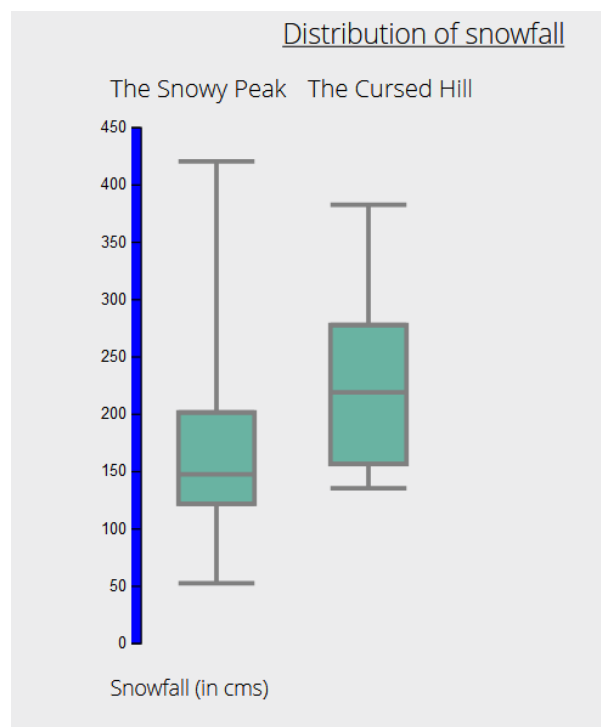
Here's our mountain 1 data...



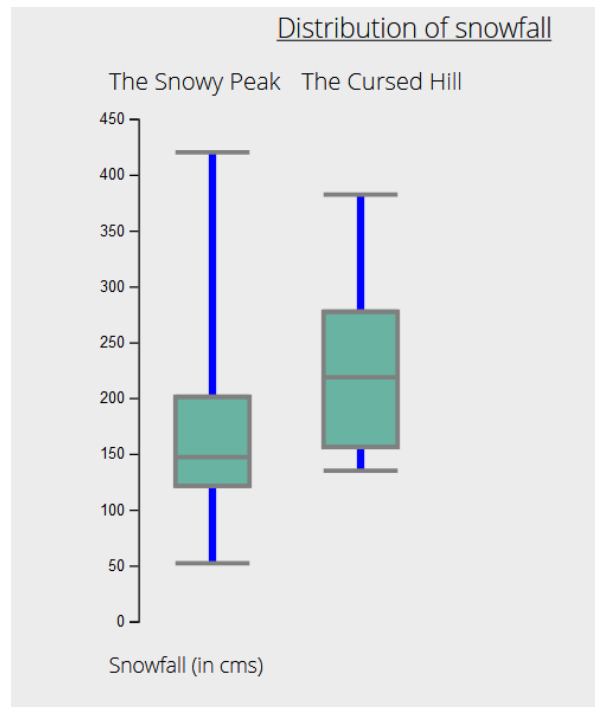
And here's some data from mountain 2.



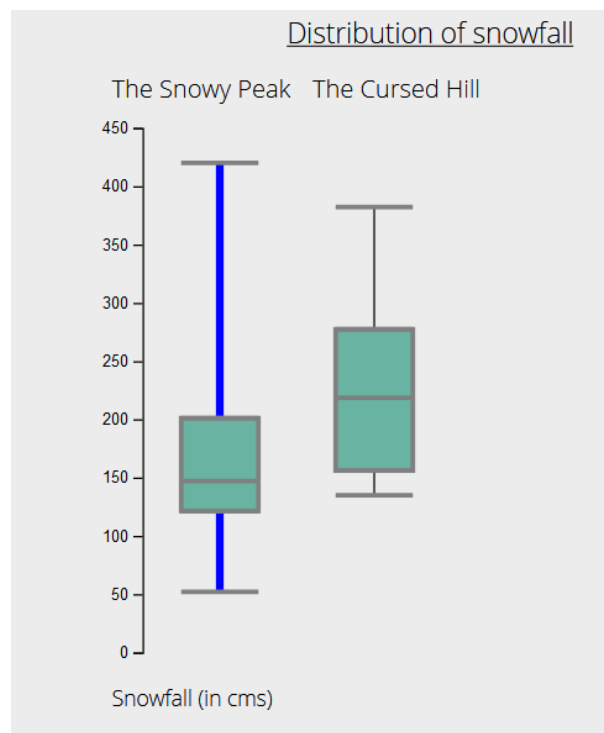
Note that the scale should be the same for them to be comparable!



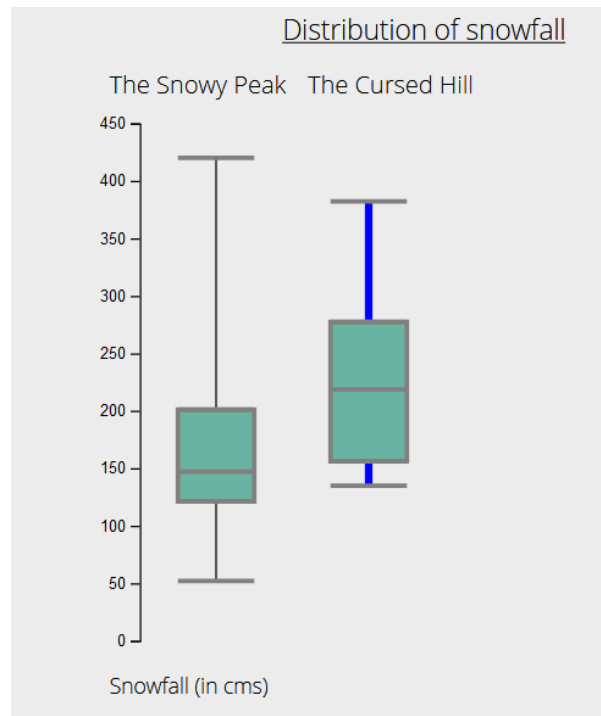
The extreme values give us information on the minimum, maximum and **variance**: the spread of data.



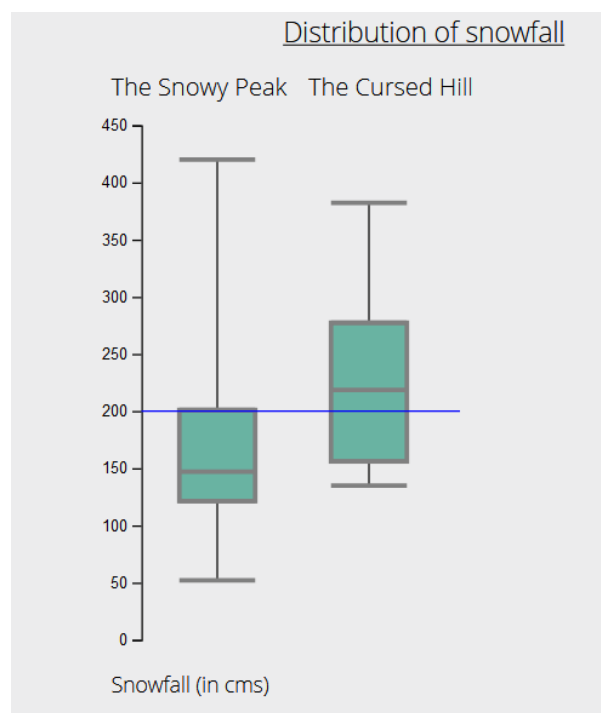
Mountain 1 has **greater variance**: this means the annual snowfall data is more spread out.



Meanwhile, Mountain 2 has **smaller variance**, which means that snowfall doesn't vary as much year to year.



We can look for the mountain with a **smaller variance**, and the **greatest chance** to receive snowfall more than a certain value, say, **200cm**.

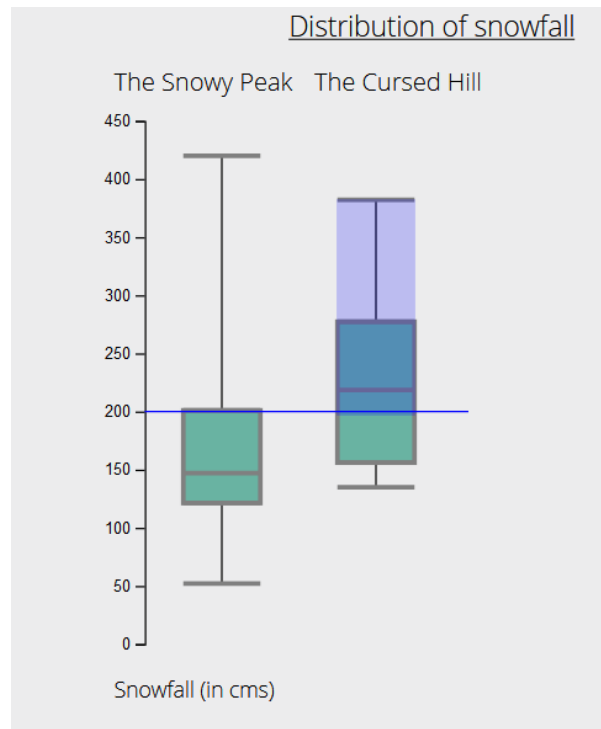


That would be mountain 2! Since its **median** is above 200cm, it gets more than 200cm of snow more than half the time.

This is the case for mountain 1 only a quarter of the time.

Amazing, we've learnt a lot. Now I can't wait for the Easter Holidays!

**The End**



Want to learn more?

To discover how to avoid making mistakes when designing your own boxplots, check out [Boxplot Pitfalls Cheatsheet](#) and [Box Plot Blunders](#).

To learn about how visualisations can be used to mislead, check out [A History Of Dishonest Fox Charts](#) and [How to Lie with Data Visualization](#).