Exploratory data analysis of temperature data

The aim of this report to provide a short summary of a basic statistical analysis of the dry bulb temperatures.

Data

The temperature data used in this analysis is from the University of Reading Atmospheric Observatory, taken at 0900 between 2009 and 2010. The temperatures are given in °C.

Methods

Exploratory data analysis was performed on the dry bulb temperatures. Basic statistical quantities are found such as the mean, standard deviation, skewness and kurtosis. Some empirical statistical quantities are also calculated. The scipy.stats and numpy packages for python have been used to calculate these quantities.

Results

The dry bulb temperature data was found to have a missing value - *you chould comment on why you think this could be*. After excluding this data it was found that the dry bulb temperature had the following statistical quantities:

Minimum(°C)	Maximum(°C)	Mean (°C)	Standard deviation	Skewness	Kurtosis
-6.099	25.0	10.24	6.54	-0.29	-0.78

You should comment on why there is such a difference between the minimum and maximum – we are looking at annual data so the minimum values are from winter and the maximums from summer – perhaps a better analysis is to look at the data seasonally. Comment on the relationship between this and the standard deviation, what do you think would happen to the standard deviation if we looked at summer and winter separately? Comment on the skewness and kurtosis, is the distribution close to Gaussian (Normal)? Perhaps looking at or showing a histogram will help you decide.

We didn't calculate the median or interquartile range (IQR) in the practical – but these are useful quantities to calculate and state in this report. In fact the upper quartile is 15.5°C and the lower quartile is 5.4°C, so 50% of the values lie in this range. What is the median?

What are the 5th and 95th percentiles? The 95th percentile is 19.5°C, so only 5% of values lie above 19.5.

Make sure you know how to calculate all these quantities! (Hint numpy.percentile will be useful)
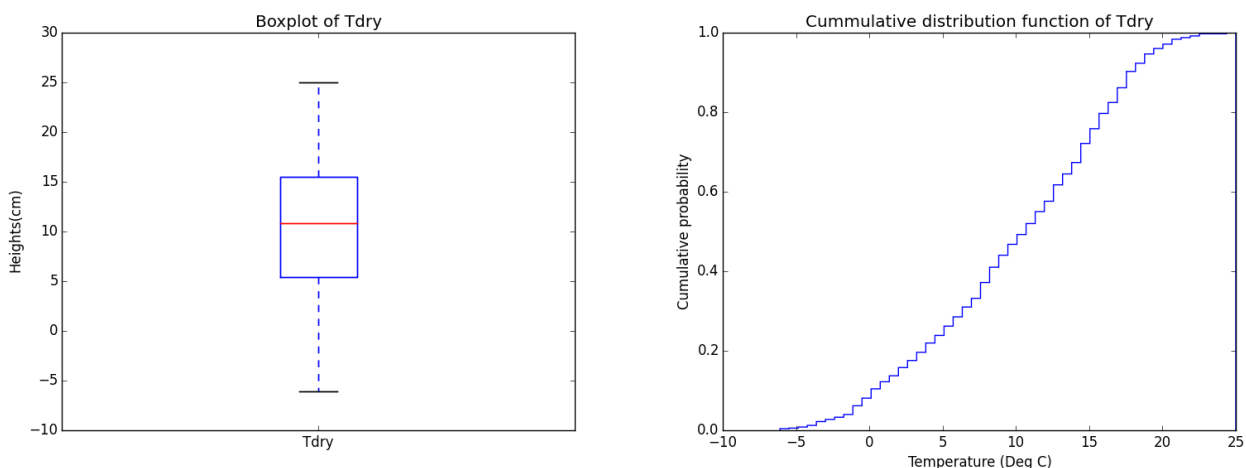

Figure 1: Boxplot of the dry bulb temperature data between 2009-2010 (left), and its cumulative distribution function with binsize=50 (right)