Analytics Foundations: Lab 1

To obtain the dataset for today's lab, execute the following code:

```
install.packages('UsingR')
library(UsingR)
data(normtemp)
```

1. Use the *normtemp* dataset to answer the following:
   a. Determine the following statistics for the variable **temperature:**

Minimum	
Maximum	
Mean	
Standard Deviation	

   b. Does **temperature** appear to be normally distributed?

   c. Create box plots for **temperature**. Are there any outliers? Display a reference line at 98.6.
   For horizontal line: `+ geom_hline(yintercept=98.6)`
   For vertical line: `+ geom_vline(xintercept=98.6)`

   Does the median body temperature seem to be 98.6 degrees?

2. Using the *Ameshousing* dataset from our in-class examples, run some distributional analysis on **Sale_Price**, **Log(Sale_Price)**, and **Gr_Liv_Area**.
   a. Create histograms of these three variables.
      - Overlay a kernel density estimator of the variables.

b. Create a QQ Plot for both **Sale_Price** and **Log(Sale_Price)**. Based on these exploratory procedures, which version of the price information (**Sale_Price** or **Log(Sale_Price)**) would you say is closer to being normally distributed?

3. Using the *Ameshousing* dataset from our in-class examples, determine the following:

   a. What type of variables are each of these columns (***Nominal, Ordinal, or Continuous/Quantitative***)? Keep in mind that the way they are represented in the R dataset may not be appropriate, so you should make this determination using your *own judgement based on the data you are looking at*.

   - **Overall_Qual**
   - **Lot_Shape**
   - **Heating_QC**
   - **Lot_Area**