Data entry - Excel

Many people use Excel, but unfortunately it is not the best mechanism for storing data long term. It poses some issues for getting the data into R for analysis (i.e., multiple sheets). The best method is to save each datasheet as a csv (comma-separated values) file. Then to input the data use:

read.csv("directory and filename", header = TRUE, sep = ",")

The header should include the variable names. The separator (sep) is normally a comma.

df <- read.csv("D:/fishdata.csv", header = TRUE, sep = ",") #Enter the data  
summary(df) #Produce a short summary of the data

## WWSurveyId\_id SurveyNum\_id FishSampNum\_id Species\_code   
## Min. :10167 Min. : 1.00 Min. : 1.00 YP :4051   
## 1st Qu.:10167 1st Qu.: 46.00 1st Qu.: 14.00 TNC : 651   
## Median :10167 Median : 80.00 Median : 35.00 PS : 348   
## Mean :10339 Mean : 75.02 Mean : 49.66 NSF : 260   
## 3rd Qu.:10167 3rd Qu.:107.00 3rd Qu.: 70.00 LMB : 248   
## Max. :16167 Max. :144.00 Max. :299.00 SMB : 223   
## (Other): 551   
## Length\_meas FishLenMo\_code Weight\_meas   
## Min. : 11.0 Mode:logical Min. : 1.0   
## 1st Qu.: 97.0 NA's:6332 1st Qu.: 11.0   
## Median :120.0 Median : 19.0   
## Mean :157.1 Mean : 137.6   
## 3rd Qu.:165.0 3rd Qu.: 52.0   
## Max. :670.0 Max. :2875.0   
## NA's :338

These data will be used for a number of subsequent explorations.