Advanced Data Analysis 1, Stat 427/527 Homework 1 Name: _____
Fall 2014 Prof. Erik B. Erhardt

Part I. (60 points) Do all calculations in R. All R code for the assignment should be included with the part of the problem it addresses (for code and output use a fixed-width font, such as Courier). Code is used to calculate result. Text is used to report and interpret results. Do not report or interpret results in the code.

(15pts) **1. Precip:** The Dept of Meteorology at the University of Stockholm monitors chemical constituents of the atmosphere at several stations throughout Sweden. The chemicals are precipitated out of the atmosphere by rain and deposited on filters, from which the amount of chemical, in milligrams per square meter of filter surface, can be measured. The monthly sulphur (Sulphur) in mg/m^2 for each of the 12 months (Month) (1=Jan, 2=Feb, etc) and the monthly precipitation (Precip) in mm is given in the table below for one station.

Month	Precip	Sulphur
1	35	55
2	25	30
3	12	25
4	36	43
5	81	135
6	19	38
7	55	63
8	63	93
9	69	64
10	23	17
11	52	34
12	35	34

Read the data from the website with:

```
d1 <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_01_F14-1.csv")
```

Refer to columns in the data using the "`$`" sign. `d1` is the dataframe, and `Precip` is a variable in the dataframe.

```
# Precip column
d1$Precip
# Sulphur column
d1$Sulphur
```

(a) (3 pts) Make a stem-and-leaf display, histogram, and boxplot for the Precip data.

(b) (2 pts) Compute the mean, median, standard deviation, and interquartile range for the Precip data.

(c) (4 pts) Is there much difference between the mean and median? Discuss, briefly, whether the size and the direction of the difference is sensible, given the graphical summaries.

(d) (6 pts) Using the graphical summaries, describe the shape of the Precip distribution. Discuss modality, presence/absence of outliers, whether skewness is present, and if so, in what direction, and whether it would be reasonable to assume that the Precip distribution is normal.

(15pts) **2. Sulphur:** Repeat with the Sulphur data in the previous problem.

(15pts) **3. Mammals:** These data, from Holling (1982), are the body masses (mass) in grams for 36 species of boreal forest mammals found east of the Manitoba-Ontario border in pure or mixed conifer strands. Repeat the steps in Problem 1 in Part I.

```
mass(g)
   3.69     104.90
```

4.11	191.36
4.25	839.15
5.24	1224.70
5.50	1496.85
8.36	2642.17
13.18	3118.45
22.45	8504.85
22.96	9071.84
24.52	10149.12
26.93	12303.68
28.30	14061.35
28.35	23995.02
33.00	43204.64
42.53	86416.53
43.94	105686.93
44.51	169643.40
80.80	481828.10

(15$^{\text{pts}}$) **4. log(Mammals):** Repeat with the natural logarithm of the data in the previous problem. You can create a new variable with something like: `logmass <-log(mass)`.