36-401 Homework #1b (Due Sept. 11, 2001) REVISED Solutions

1. **Create a random vector using "ch_rchisq(40,1)". Plot a histogram, choosing a number of bars that displays the data well. Turn in your plot and the mean and median of the vector.**
   > ch_rchisq(40,1)
   > hist(ch, nclass=12)

   The output is displayed in Figure 1.

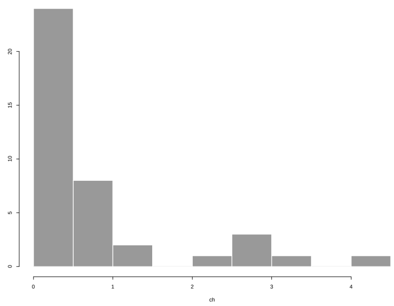

Figure 1: Histogram of 40 random chi-square (1 df) numbers

> summary(ch)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0002709	0.05135	0.3097	0.7127	0.9237	4.06

Therefore, the mean is .7127 and the median is .3097.

2. **Load the data in the file "widgets.dat". Turn in the list of variable names, the number of rows, and the range (max-min) of each column.**
   > widge_read.table("widgets.dat", header=T)
   > names(widge)

[1] "whipples" "waffles" "woozles"

> summary(widge)

whipples	waffles	woozles
Min. : 9.868	Min. :-9.9800	Min. :0.00
1st Qu.:16.310	1st Qu.:-4.7620	1st Qu.:0.75
Median :20.710	Median :-0.1579	Median :1.00
Mean :20.260	Mean : 0.1369	Mean :0.75
3rd Qu.:23.590	3rd Qu.: 5.5570	3rd Qu.:1.00
Max. :34.030	Max. : 9.7740	Max. :1.00

> dim(widge)

[1] 100 3

Therefore, there are 100 rows (and 3 columns).

3. **Download "cane.dat" and "cane.txt" form the class website. Download the ndhist.q, rwgbox.q, symplot.q, qplot.q, qqn.q, and scatbox.q from the website. Check that you have all of the files using the Unix command "ls". Print "cane.txt" or read it with th eUnix "more" command. Use read.table() to load the data in the "cane.dat" file into an object named "cane." Check the dim(cane) and names(cane). View columns 6 to 12 for the first 10 cass using "cane[1:10,6:12]". Load each of the functions in the .q files listed above using the source function. Try the first 5 functions on the sugar variable, e.g. ndhist(cane$Sugar). Also run "scatbox(cane$Sugar, cane$Area)" to compare these variables. Use the "mfrow=" argument of the par() command to create a single graph containing all 6 subgraphs. Print it, including your name as a subtitle.**

```
> motif()
> source("ndhist.q")
> source("rwgbox.q")
> source("symplot.q")
> source("qplot.q")
> source("qqn.q")
> source("scatbox.q")
> cane_read.table("cane.dat", header=T)
> dim(cane)
```

[1] 3775 32

> names(cane)

[1]	"District"	"DistrictGroup"	"DistrictPosition"	"SoilID"
[5]	"SoilName"	"Area"	"Variety"	"Ratoon"
[9]	"Age"	"HarvestMonth"	"HarvestDuration"	"TonnPerHect"
[13]	"Fibre"	"Sugar"	"Jul96"	"Aug96"
[17]	"Sep96"	"Oct96"	"Nov96"	"Dec96"
[21]	"Jan97"	"Feb97"	"Mar97"	"Apr97"
[25]	"May97"	"Jun97"	"Jul97"	"Aug97"
[29]	"Sep97"	"Oct97"	"Nov97"	"Dec97"

> cane[1:10, 6:12]

	Area	Variety	Ratoon	Age	HarvestMonth	HarvestDuration	TonnPerHect
1	0.81	152	1R	1	6	3	47.09
2	9.87	124	1R	1	8	55	1243.41
3	1.53	158	2R	2	9	0	141.56
4	3.79	138	2R	2	9	0	310.69
5	2.81	124	2R	2	9	1	219.02
6	6.44	124	PL	0	7	5	902.46
7	1.04	124	PL	0	7	0	154.55
8	1.04	124	PL	0	7	0	128.38
9	7.03	120	3R	3	8	1	553.99
10	1.27	120	2R	2	9	0	86.47

My program, plots.q, creates the six requested plots and includes my name as a subtitle at the bottom of the page:

```
plots_function(x, y)

{ source("ndhist.q")
source("rwgbox.q")
source("symplot.q")
source("qplot.q")
source("qqn.q")
source("scatbox.q")
par(mfrow=c(3,2), oma=c(3,1,1,1))
ndhist(x)
rwgbox(x)
symplot(x)
qplot(x)
qqn(x)
scatbox(x, y)
mtext("Kimberly Sellers",cex=1.2,outer=T,side=1)
}
```

Thus, the following commands are inputted in Splus: > motif()

```
> input_read.table("cane.dat", header=T)
> source("plots.q")
> plots(cane$Sugar, cane$Area)
```

and the output results in Figure 2.

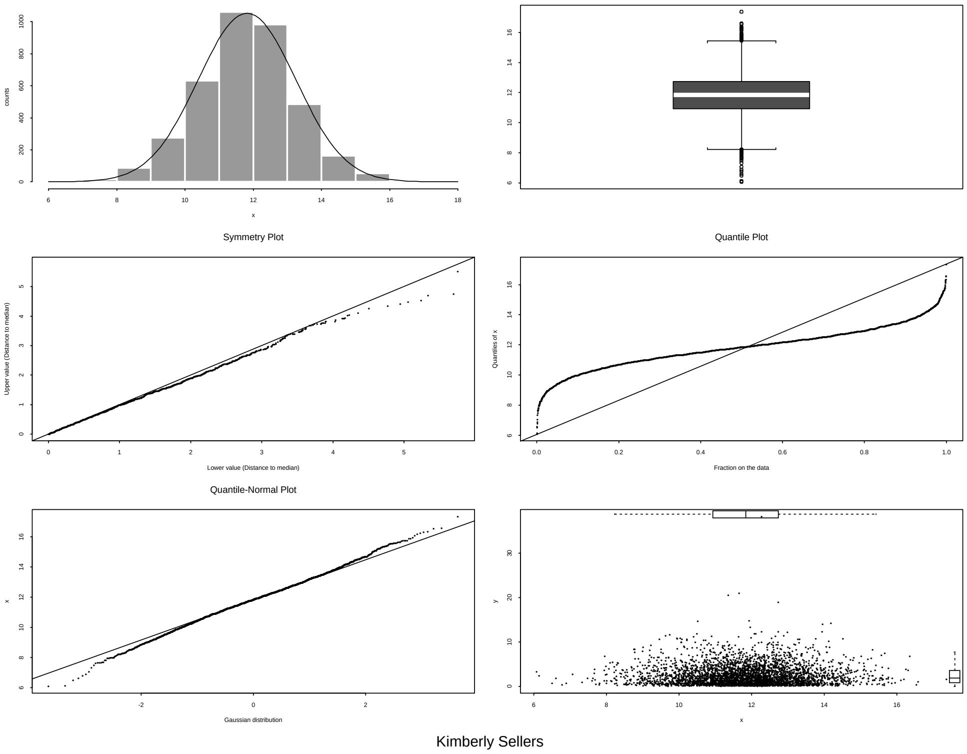

Figure 2: Plots of sugar content per rake and sugar content vs. area of paddock