STAT 400 Sept. 9, 2005

Some Details:

Class MWF 1130-1220 K 9500.
Office Hours for KLW: M 1330 and F 1230 (but not Sept 9).
SSC K 10565. 604- 291-3667. weldon@sfu.ca
Notes at www.stat.sfu.ca/~weldon
TA is Darcy Pickard dcp@sfu.ca
SSC K 9501 (604) 291-4983
(no tutorial Monday Sept 12).

Today:

Review of Ch 1:

Barley Data portrayed on p 4:

	yield	variety	year	site
1	27.00000	Manchuria	1931	University Farm
2	48.86667	Manchuria	1931	Waseca
3	27.43334	Manchuria	1931	Morris
4	39.93333	Manchuria	1931	Crookston
5	32.96667	Manchuria	1931	Grand Rapids
6	28.96667	Manchuria	1931	Duluth
7	43.06666	Glabron	1931	University Farm
8	55.20000	Glabron	1931	Waseca
9	28.76667	Glabron	1931	Morris
10	38.13333	Glabron	1931	Crookston
11	29.13333	Glabron	1931	Grand Rapids
12	29.66667	Glabron	1931	Duluth
13	35.13333	Svansota	1931	University Farm
14	47.33333	Svansota	1931	Waseca
15	25.76667	Svansota	1931	Morris
16	40.46667	Svansota	1931	Crookston
17	29.66667	Svansota	1931	Grand Rapids
18	25.70000	Svansota	1931	Duluth
19	39.90000	Velvet	1931	University Farm
20	50.23333	Velvet	1931	Waseca
21	26.13333	Velvet	1931	Morris
22	41.33333	Velvet	1931	Crookston
23	23.03333	Velvet	1931	Grand Rapids
24	26.30000	Velvet	1931	Duluth

25	36.56666	Trebi	1931	University Farm
26	63.83330	Trebi	1931	Waseca
27	43.76667	Trebi	1931	Morris
28	46.93333	Trebi	1931	Crookston
29	29.76667	Trebi	1931	Grand Rapids
30	33.93333	Trebi	1931	Duluth
31	43.26667	No. 457	1931	University Farm
32	58.10000	No. 457	1931	Waseca
33	28.70000	No. 457	1931	Morris
34	45.66667	No. 457	1931	Crookston
35	32.16667	No. 457	1931	Grand Rapids
36	33.60000	No. 457	1931	Duluth
37	36.60000	No. 462	1931	University Farm
38	65.76670	No. 462	1931	Waseca
39	30.36667	No. 462	1931	Morris
40	48.56666	No. 462	1931	Crookston
41	24.93334	No. 462	1931	Grand Rapids
42	28.10000	No. 462	1931	Duluth
43	32.76667	Peatland	1931	University Farm
44	48.56666	Peatland	1931	Waseca
45	29.86667	Peatland	1931	Morris
46	41.60000	Peatland	1931	Crookston
47	34.70000	Peatland	1931	Grand Rapids
48	32.00000	Peatland	1931	Duluth
49	24.66667	No. 475	1931	University Farm
50	46.76667	No. 475	1931	Waseca
51	22.60000	No. 475	1931	Morris
52	44.10000	No. 475	1931	Crookston
53	19.70000	No. 475	1931	Grand Rapids
54	33.06666	No. 475	1931	Duluth
55	39.30000	Wisconsin No. 38	1931	University Farm
56	58.80000	Wisconsin No. 38	1931	Waseca
57	29.46667	Wisconsin No. 38	1931	Morris
58	49.86667	Wisconsin No. 38	1931	Crookston
59	34.46667	Wisconsin No. 38	1931	Grand Rapids
60	31.60000	Wisconsin No. 38	1931	Duluth
61	26.90000	Manchuria	1932	University Farm
62	33.46667	Manchuria	1932	Waseca
63	34.36666	Manchuria	1932	Morris
64	32.96667	Manchuria	1932	Crookston
65	22.13333	Manchuria	1932	Grand Rapids
66	22.56667	Manchuria	1932	Duluth
67	36.80000	Glabron	1932	University Farm
68	37.73333	Glabron	1932	Waseca
69	35.13333	Glabron	1932	Morris
70	26.16667	Glabron	1932	Crookston
71	14.43333	Glabron	1932	Grand Rapids
72	25.86667	Glabron	1932	Duluth

73	27.43334	Svansota	1932	University Farm
74	38.50000	Svansota	1932	Waseca
75	35.03333	Svansota	1932	Morris
76	20.63333	Svansota	1932	Crookston
77	16.63333	Svansota	1932	Grand Rapids
78	22.23333	Svansota	1932	Duluth
79	26.80000	Velvet	1932	University Farm
80	37.40000	Velvet	1932	Waseca
81	38.83333	Velvet	1932	Morris
82	32.06666	Velvet	1932	Crookston
83	32.23333	Velvet	1932	Grand Rapids
84	22.46667	Velvet	1932	Duluth
85	29.06667	Trebi	1932	University Farm
86	49.23330	Trebi	1932	Waseca
87	46.63333	Trebi	1932	Morris
88	41.83333	Trebi	1932	Crookston
89	20.63333	Trebi	1932	Grand Rapids
90	30.60000	Trebi	1932	Duluth
91	26.43334	No. 457	1932	University Farm
92	42.20000	No. 457	1932	Waseca
93	43.53334	No. 457	1932	Morris
94	34.33333	No. 457	1932	Crookston
95	19.46667	No. 457	1932	Grand Rapids
96	22.70000	No. 457	1932	Duluth
97	25.56667	No. 462	1932	University Farm
98	44.70000	No. 462	1932	Waseca
99	47.00000	No. 462	1932	Morris
100	30.53333	No. 462	1932	Crookston
101	19.90000	No. 462	1932	Grand Rapids
102	22.50000	No. 462	1932	Duluth
103	28.06667	Peatland	1932	University Farm
104	36.03333	Peatland	1932	Waseca
105	43.20000	Peatland	1932	Morris
106	25.23333	Peatland	1932	Crookston
107	26.76667	Peatland	1932	Grand Rapids
108	31.36667	Peatland	1932	Duluth
109	30.00000	No. 475	1932	University Farm
110	41.26667	No. 475	1932	Waseca
111	44.23333	No. 475	1932	Morris
112	32.13333	No. 475	1932	Crookston
113	15.23333	No. 475	1932	Grand Rapids
114	27.36667	No. 475	1932	Duluth
115	38.00000	Wisconsin No. 38	1932	University Farm
116	58.16667	Wisconsin No. 38	1932	Waseca
117	47.16667	Wisconsin No. 38	1932	Morris
118	35.90000	Wisconsin No. 38	1932	Crookston
119	20.66667	Wisconsin No. 38	1932	Grand Rapids
120	29.33333	Wisconsin No. 38	1932	Duluth

Exercise:

See if you can summarize the barley data using any methods/software available to you.

Report your experience on Friday, Sept 16. (Class discussion).

Some questions arising in Chapter 1:

1. What is wrong with Fig 1.2 (singers)?
2. In Figure 1.3 (polarization), what role would be fulfilled by fitting a curve to the data? Can residuals be computed without a parametric fit?
3. Fig. 1.5: Can a 3-variable data structure be gleaned from 2-variable plots? (NOx)
4. What does Cleveland mean by "Direct Manipulation"?

Ch 2: Univariate Data

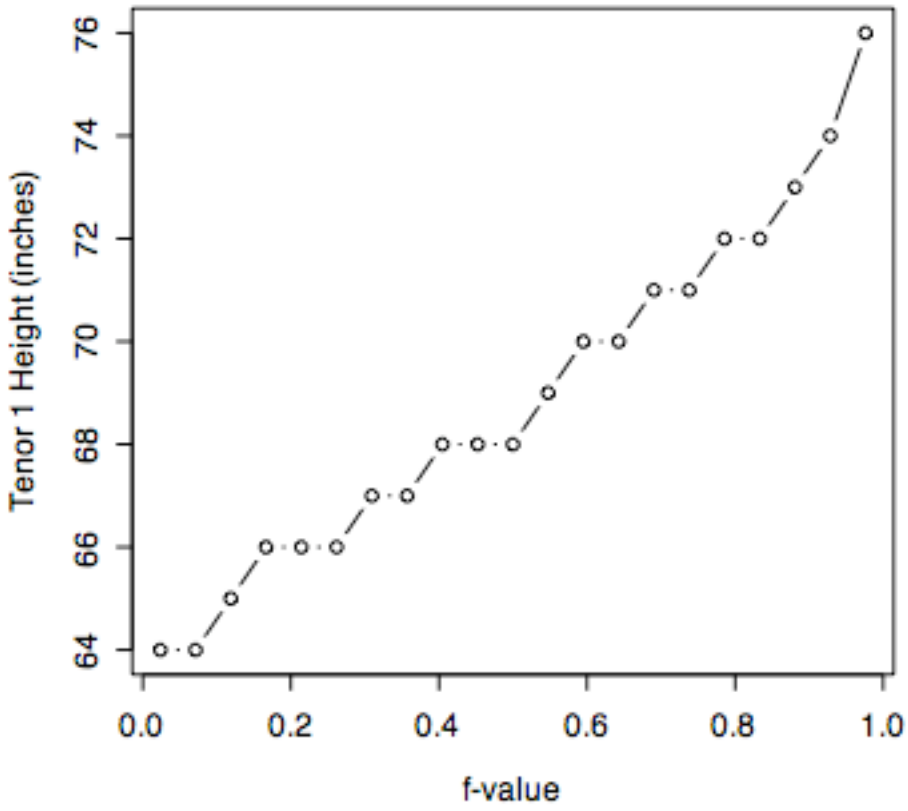
Read Ch 2 pp 17-46 for Monday.

Quantiles:

```
read.table("tenor.txt")
    V1 V2    V3 V4
1  129 69 Tenor  1
2  130 72 Tenor  1
3  131 71 Tenor  1
4  132 66 Tenor  1
5  133 76 Tenor  1
6  134 74 Tenor  1
7  135 71 Tenor  1
8  136 66 Tenor  1
9  137 68 Tenor  1
10 138 67 Tenor  1
11 139 70 Tenor  1
12 140 65 Tenor  1
```

```
13 141 72 Tenor  1
14 142 70 Tenor  1
15 143 68 Tenor  1
16 144 64 Tenor  1
17 145 73 Tenor  1
18 146 66 Tenor  1
19 147 68 Tenor  1
20 148 67 Tenor  1
21 149 64 Tenor  1
> a=read.table("tenor.txt")
> tenor.df=as.data.frame(a$V2)
> colnames(tenor.df)=c("height")
> tenor.df
   height
1      69
2      72
3      71
4      66
5      76
6      74
7      71
8      66
9      68
10     67
11     70
12     65
13     72
14     70
15     68
16     64
17     73
18     66
19     68
20     67
21     64
> a=sort(tenor.df$height)
```

```
> a
 [1] 64 64 65 66 66 66 67 67 68 68 68 69 70 70 71 71 72 72 73 74 76
> b=(1:210.5)/21
> plot(b,a,type="l")
> plot(b,a,type="b",xlab="fvalue",ylab="Tenor 1 Height (inches)")
```



What is the f=.2 quantile of the height distribution? Ans = 66.

Roughly The f-quantile is q(f) = the value that exceeds a proportion f of the distribution. See p 1819.

Note that quantile comparisons between two distributions are not systematically affected by the sample size. See Fig 2.1 p 16.

QQ plots p 21

With two variables, just plot as a scatter plot of (x,y) the x and y that correspond to approximately the same quantile. If the sample sizes are unequal, find the number of (x,y) values equal to the smaller sample size.

How would Fig 2.3 compare with a scatter plot of bass2 vs tenor1?

Important to understand this to appreciate the QQ plot!