Math 3070 § 1. **Particulate Data:** Name: Example
Treibergs **Sample Statistics and Boxplots.** September 4, 2010

Data File Used in this Analysis:

```
# Math 3070 - 1     Particulate Data        Treibergs
#
# From Navidi, "Principles of Statistics for Engineers and Scientists"
# McGraw Hill, 2010. This data is taken from a study by J. Yanowitz,
# Colorado School of Mines. PM is automobile emissions (in g/gal)
# for a sample of 138 cars driven at low altitude vs 62 cars driven
# at high altitude (one mile above sea level) which were manufactured
# between 1991 and 1996.
#
PM Altitude
1.5     Low
0.87     Low
1.12     Low
1.25     Low
3.46     Low
1.11     Low
1.12     Low
0.88     Low
1.29     Low
0.94     Low
0.64     Low
1.31     Low
2.49     Low
1.48     Low
1.06     Low
1.11     Low
2.15     Low
0.86     Low
1.81     Low
1.47     Low
1.24     Low
1.63     Low
2.14     Low
6.64     Low
4.04     Low
2.48     Low
2.98     Low
7.39     Low
2.66     Low
11      Low
4.57     Low
4.38     Low
0.87     Low
1.1     Low
1.11     Low
0.61     Low
1.46     Low
```

0.97	Low
0.9	Low
1.4	Low
1.37	Low
1.81	Low
1.14	Low
1.63	Low
3.67	Low
0.55	Low
2.67	Low
2.63	Low
3.03	Low
1.23	Low
1.04	Low
1.63	Low
3.12	Low
2.37	Low
2.12	Low
2.68	Low
1.17	Low
3.34	Low
3.79	Low
1.28	Low
2.1	Low
6.55	Low
1.18	Low
3.06	Low
0.48	Low
0.25	Low
0.53	Low
3.36	Low
3.47	Low
2.74	Low
1.88	Low
5.94	Low
4.24	Low
3.52	Low
3.59	Low
3.1	Low
3.33	Low
4.58	Low
6.73	Low
7.82	Low
4.59	Low
5.12	Low
5.67	Low
4.07	Low
4.01	Low
2.72	Low
3.24	Low
5.79	Low
3.59	Low

```
3.48     Low
2.96     Low
5.3     Low
3.93     Low
3.52     Low
2.96     Low
3.12     Low
1.07     Low
5.3     Low
5.16     Low
7.74     Low
5.41     Low
3.4     Low
4.97     Low
11.23   Low
9.3     Low
6.5     Low
4.62     Low
5.45     Low
4.93     Low
6.05     Low
5.82     Low
10.19   Low
3.62     Low
2.67     Low
2.75     Low
8.92     Low
9.93     Low
6.96     Low
5.78     Low
9.14     Low
10.63   Low
8.23     Low
6.83     Low
5.6     Low
5.41     Low
6.7     Low
5.93     Low
4.51     Low
9.04     Low
7.71     Low
7.21     Low
4.67     Low
4.49     Low
4.63     Low
2.8     Low
2.16     Low
2.97     Low
3.9     Low
7.59     High
6.28     High
6.07     High
```

```
5.23     High
5.54     High
3.46     High
2.44     High
3.01     High
13.63    High
13.02   High
23.38   High
9.24     High
3.22     High
2.06     High
4.04     High
17.11   High
12.26   High
19.91   High
8.5     High
7.81     High
7.18     High
6.95     High
18.64   High
7.1     High
6.04     High
5.66     High
8.86     High
4.4     High
3.57     High
4.35     High
3.84     High
2.37     High
3.81     High
5.32     High
5.84     High
2.89     High
4.68     High
1.85     High
9.14     High
8.67     High
9.52     High
2.68     High
10.14   High
9.2     High
7.31     High
2.09     High
6.32     High
6.53     High
6.32     High
2.01     High
5.91     High
5.6     High
5.61     High
1.5     High
6.46     High
```

```
5.29     High
5.64     High
2.07     High
1.11     High
3.32     High
1.83     High
7.56     High
```

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5537) powerpc-apple-darwin9.8.0]

> tt <- read.table("M3073ParticleData.txt",header=TRUE)
Error in file(file, "rt") : cannot open the connection
In addition: Warning message:
In file(file, "rt") :
  cannot open file 'M3073ParticleData.txt': No such file or directory
> tt <- read.table("M3070ParticulateData.txt",header=TRUE)
> tt
       PM Altitude
1    1.50      Low
2    0.87      Low
3    1.12      Low
4    1.25      Low
5    3.46      Low
6    1.11      Low
7    1.12      Low
8    0.88      Low
9    1.29      Low
10   0.94      Low
11   0.64      Low
12   1.31      Low
13   2.49      Low
```

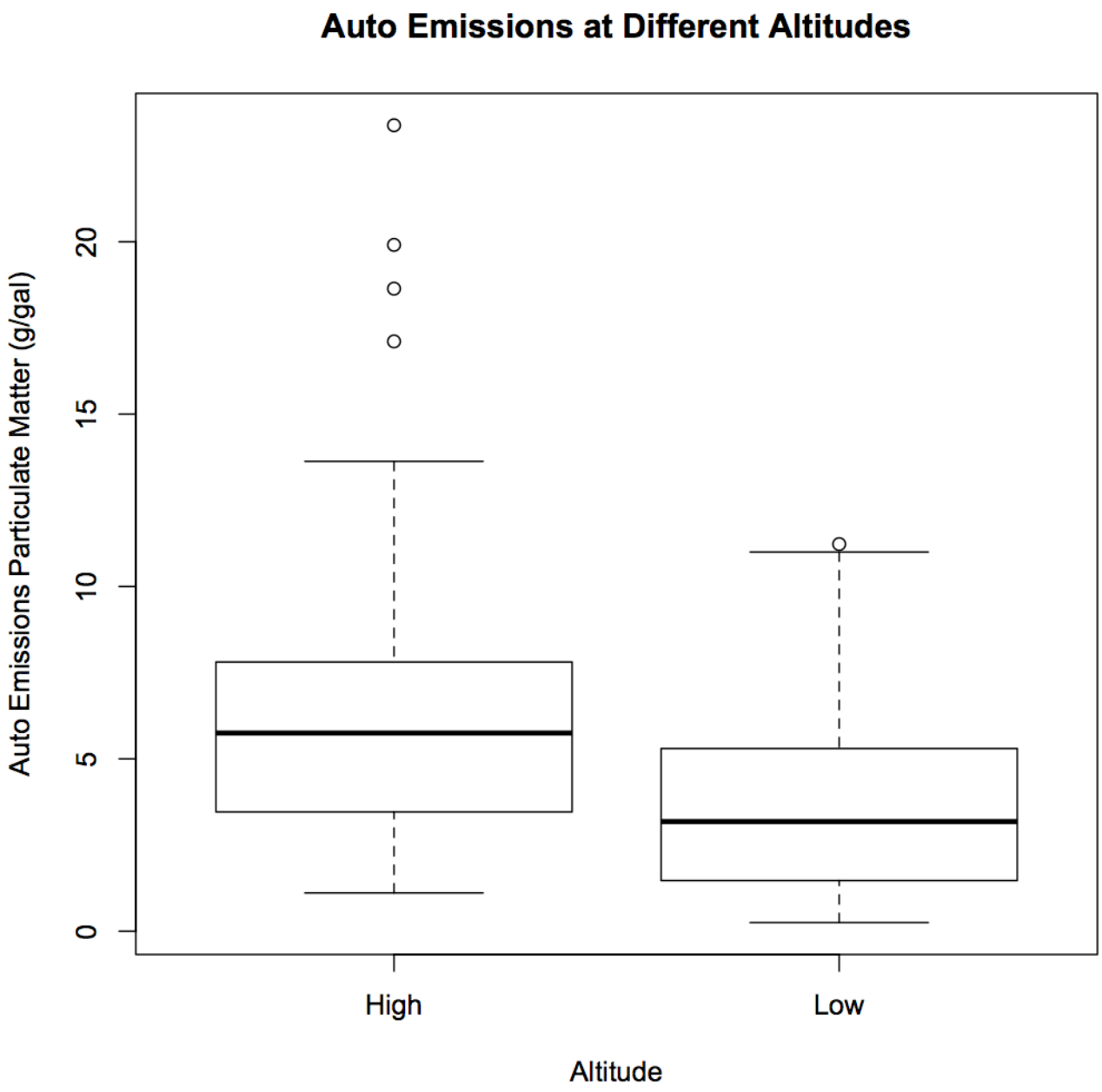
14	1.48	Low
15	1.06	Low
16	1.11	Low
17	2.15	Low
18	0.86	Low
19	1.81	Low
20	1.47	Low
21	1.24	Low
22	1.63	Low
23	2.14	Low
24	6.64	Low
25	4.04	Low
26	2.48	Low
27	2.98	Low
28	7.39	Low
29	2.66	Low
30	11.00	Low
31	4.57	Low
32	4.38	Low
33	0.87	Low
34	1.10	Low
35	1.11	Low
36	0.61	Low
37	1.46	Low
38	0.97	Low
39	0.90	Low
40	1.40	Low
41	1.37	Low
42	1.81	Low
43	1.14	Low
44	1.63	Low
45	3.67	Low
46	0.55	Low
47	2.67	Low
48	2.63	Low
49	3.03	Low
50	1.23	Low
51	1.04	Low
52	1.63	Low
53	3.12	Low
54	2.37	Low
55	2.12	Low
56	2.68	Low
57	1.17	Low
58	3.34	Low
59	3.79	Low
60	1.28	Low
61	2.10	Low
62	6.55	Low
63	1.18	Low
64	3.06	Low
65	0.48	Low

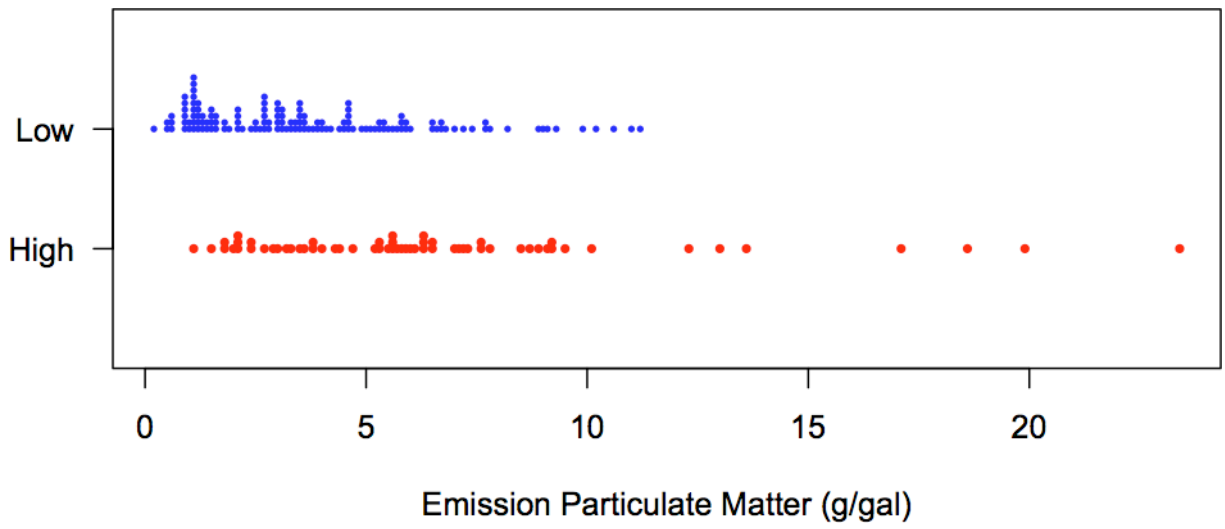
66	0.25	Low
67	0.53	Low
68	3.36	Low
69	3.47	Low
70	2.74	Low
71	1.88	Low
72	5.94	Low
73	4.24	Low
74	3.52	Low
75	3.59	Low
76	3.10	Low
77	3.33	Low
78	4.58	Low
79	6.73	Low
80	7.82	Low
81	4.59	Low
82	5.12	Low
83	5.67	Low
84	4.07	Low
85	4.01	Low
86	2.72	Low
87	3.24	Low
88	5.79	Low
89	3.59	Low
90	3.48	Low
91	2.96	Low
92	5.30	Low
93	3.93	Low
94	3.52	Low
95	2.96	Low
96	3.12	Low
97	1.07	Low
98	5.30	Low
99	5.16	Low
100	7.74	Low
101	5.41	Low
102	3.40	Low
103	4.97	Low
104	11.23	Low
105	9.30	Low
106	6.50	Low
107	4.62	Low
108	5.45	Low
109	4.93	Low
110	6.05	Low
111	5.82	Low
112	10.19	Low
113	3.62	Low
114	2.67	Low
115	2.75	Low
116	8.92	Low
117	9.93	Low

```
118  6.96      Low
119  5.78      Low
120  9.14      Low
121 10.63      Low
122  8.23      Low
123  6.83      Low
124  5.60      Low
125  5.41      Low
126  6.70      Low
127  5.93      Low
128  4.51      Low
129  9.04      Low
130  7.71      Low
131  7.21      Low
132  4.67      Low
133  4.49      Low
134  4.63      Low
135  2.80      Low
136  2.16      Low
137  2.97      Low
138  3.90      Low
139  7.59     High
140  6.28     High
141  6.07     High
142  5.23     High
143  5.54     High
144  3.46     High
145  2.44     High
146  3.01     High
147 13.63     High
148 13.02     High
149 23.38     High
150  9.24     High
151  3.22     High
152  2.06     High
153  4.04     High
154 17.11     High
155 12.26     High
156 19.91     High
157  8.50     High
158  7.81     High
159  7.18     High
160  6.95     High
161 18.64     High
162  7.10     High
163  6.04     High
164  5.66     High
165  8.86     High
166  4.40     High
167  3.57     High
168  4.35     High
169  3.84     High
```

```
170  2.37     High
171  3.81     High
172  5.32     High
173  5.84     High
174  2.89     High
175  4.68     High
176  1.85     High
177  9.14     High
178  8.67     High
179  9.52     High
180  2.68     High
181 10.14     High
182  9.20     High
183  7.31     High
184  2.09     High
185  6.32     High
186  6.53     High
187  6.32     High
188  2.01     High
189  5.91     High
190  5.60     High
191  5.61     High
192  1.50     High
193  6.46     High
194  5.29     High
195  5.64     High
196  2.07     High
197  1.11     High
198  3.32     High
199  1.83     High
200  7.56     High
> attach(tt)
> # The second variable is a "factor" that determines whether the
> $ observation is from "Low" or "High" altitude.
> Alt <- factor(Altitude)


> # The plot of a single variable for each factor level generates a
> # boxplot. Otherwise, use the instruction boxplot.
> plot(PM ~ Alt,ylab="Auto Emissions Particulate Matter (g/gal)",
  xlab="Altitude", main="Auto Emissions at Different Altitudes")

> # We make the dotplots of the two emissions data sets.
> layout(matrix(1:2,ncol=1))
> stripchart(round(PM,1) ~ Alt, method="stack", at=c(1.5,2), col=c(2,4),
  cex=.5, pch=19:20,xlab="Emission Particulate Matter (g/gal)")
>
```

```
> # The read.table produces a data-frame. Summary produces min/max,
> # quartiles and mean for variable in the frame.
> summary(tt)
       PM         Altitude
 Min.   : 0.250   High: 62
 1st Qu.: 2.085   Low :138
 Median : 3.730
 Mean   : 4.608
 3rd Qu.: 6.055
 Max.   :23.380


> # The summary applied to a single vector produces the six numbers for it.
> # tapply separates the vector PM according to levels of the factor Alt
> tapply(PM, Alt, summary)
$High
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.110   3.488   5.750   6.596   7.755  23.380

$Low
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.250   1.473   3.180   3.715   5.265  11.230

> # here i define the function allstat. For the vector x it makes a little
> # table with n, mean, sd and var including labels.
> allstat <- function(x){matrix(c(length(x),mean(x),sd(x),var(x)),  ncol=4,
  dimnames=list("",
  c("n","  Sample Mean","  Sample Std. Dev.","  Sample Variance")))}
> tapply(PM,Alt,allstat)
$High
  n   Sample Mean   Sample Std. Dev.   Sample Variance
 62      6.596452           4.518998          20.42134

$Low
   n   Sample Mean   Sample Std. Dev.   Sample Variance
 138      3.714565            2.55804           6.54357

> # Finally, we print the two dotplots on the same horizontal scale
> xr <- range(PM)*c(0.9,1.1)
> layout(matrix(1:2,ncol=1))
> stripchart(round(PM,1) ~ Alt, method="stack", at=c(1.5,2), col=c(2,4),
  cex=.5, pch=19:20, xlab="Emission Particulate Matter (g/gal)")
```


Auto Emissions at Different Altitudes
Auto Emissions Particulate Matter (g/gal)
0
5
10
15
20
High
Low
Altitude


Low
High
0
5
10
15
20
Emission Particulate Matter (g/gal)