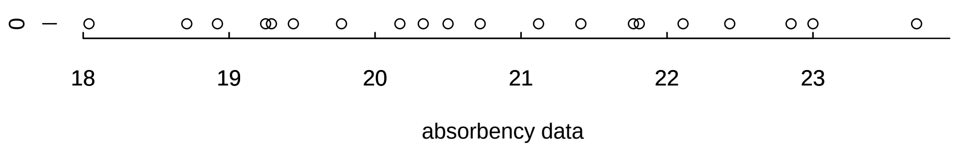
1. Exercise 1.2 in text. Show all calculations by hand.

   - Solution:

     The sample mean is 20.77 and the sample median is 20.61, which is the average of the two middle points 20.5 and 20.72. The 10% trimmed mean is the average of the middle 16 values, found by removing observations $18.04, 18.71, 23, 23.71$. It is 20.74. There is no evidence of outliers using the above descriptive statistics. The mean and median are close, and the mean is barely affected by trimming. Below is a dot plot.



2. Exercise 1.9 in text. Show all calculations by hand.

   - Solution:

     The variance and standard deviation of the no-aging group is 23.66 and 4.86. The variance and standard deviation of the aging group is 42.1 and 6.49. You may calculate variance using either of the two equivalent formulas:

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

or

$$\frac{n}{n-1}[(\frac{1}{n}\sum_{i=1}^{n}x_i^2) - \bar{x}]$$

where n is the number of data points and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean. Since the variance of the aging group is much higher than the no aging group, we have evidence that aging affects the variability of tensile strength.

3. `R` guided exercise. Print and hand in all `R` console output along with written answers to questions.

   a) You will complete the following guided exercise using the `R` software. You may access the `R` program in the following university labs: SPHPI 178, SPHPI B34, EMTSWL78, DH170, Library1010, and SPHW 523 labs. Alternatively (and probably preferably), you may download `R` for free at `https : //cran.r − project.org/`.

   b) Open `R` and run the following code. Remember that you may run code in two ways: by copying and pasting it into the console or by opening a new script file, copying and pasting it and then highlighting and running it with `ctrl` + `r`.

      i. `R` comes with some built-in example datasets. The following is an `R` data frame containing 6 types of bugspray and several trials of each. You may interpret the "count" variable as the efficacy of the spray in a given trial.

```
InsectSprays
```

ii. Data frames can be easily visualized in R. Here is a nice dot plot. From the dot plot, which sprays seem to be effective and which seem less so?

```
plot(InsectSprays, col = c(1), pch = 20)
```

iii. We can calculate several descriptive statistics in R. To consider just the data for spray 'A', we use the "subset" function. Then, for this subset, we apply functions to calculate various statistics. When you copy/paste the codes below, you may have to alter the quotes in 'A'. After you run the codes below, try to calculate the sample mean for each other spray (B,C,D,E) by modifying the codes below. Which spray has the highest mean? Turn in the results of your codes by printing the contents of the R console and make sure you have answered all of my questions.

```
summary(subset(InsectSprays,spray == 'A'))

median(subset(InsectSprays,spray == 'A')$count)

mean(subset(InsectSprays,spray == 'A')$count)

IQR(subset(InsectSprays,spray == 'A')$count)

mean(subset(InsectSprays,spray == 'A')$count, trim = 1/12)

var(subset(InsectSprays,spray == 'A')$count)
```

```
sd(subset(InsectSprays,spray == 'A')$count)
```

- Solutions:

  Running the first line of code, we get the following data output:

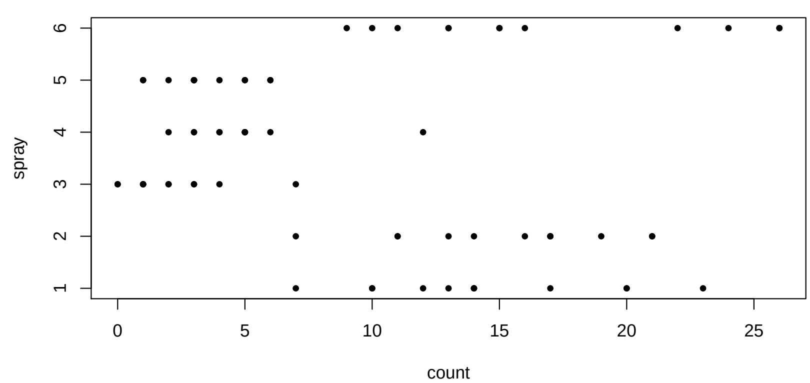
```
> InsectSprays
   count spray
1     10     A
2      7     A
3     20     A
4     14     A
5     14     A
6     12     A
7     10     A
8     23     A
9     17     A
10    20     A
11    14     A
12    13     A
13    11     B
14    17     B
15    21     B
16    11     B
17    16     B
```

18	14	B
19	17	B
20	17	B
21	19	B
22	21	B
23	7	B
24	13	B
25	0	C
26	1	C
27	7	C
28	2	C
29	3	C
30	1	C
31	2	C
32	1	C
33	3	C
34	0	C
35	1	C
36	4	C
37	3	D
38	5	D
39	12	D
40	6	D
41	4	D

42	3	D
43	5	D
44	5	D
45	5	D
46	5	D
47	2	D
48	4	D
49	3	E
50	5	E
51	3	E
52	5	E
53	3	E
54	6	E
55	1	E
56	1	E
57	3	E
58	2	E
59	6	E
60	4	E
61	11	F
62	9	F
63	15	F
64	22	F
65	15	F

```
66    16    F
67    13    F
68    10    F
69    26    F
70    26    F
71    24    F
72    13    F
```

Running the code for the plot we get the picture below.



It looks like sprays 1, 2, and 6 work well, while sprays 3, 4, and 5 work less well.

Running the descriptive statistics we get

```
> summary(subset(InsectSprays,spray == 'A'))
     count        spray
 Min.   : 7.00   A:12
 1st Qu.:11.50   B: 0
```

```
 Median :14.00   C: 0
 Mean   :14.50   D: 0
 3rd Qu.:17.75   E: 0
 Max.   :23.00   F: 0
>
> median(subset(InsectSprays,spray == 'A')$count)
[1] 14
>
> mean(subset(InsectSprays,spray == 'A')$count)
[1] 14.5
>
> IQR(subset(InsectSprays,spray == 'A')$count)
[1] 6.25
>
> mean(subset(InsectSprays,spray == 'A')$count, trim = 1/12)
[1] 14.4
>
> var(subset(InsectSprays,spray == 'A')$count)
[1] 22.27273
>
> sd(subset(InsectSprays,spray == 'A')$count)
[1] 4.719399
```

We can modify the code to calculate the other means. Here's how we do that:

```
> mean(subset(InsectSprays,spray == 'B')$count)
[1] 15.33333
> mean(subset(InsectSprays,spray == 'C')$count)
[1] 2.083333
> mean(subset(InsectSprays,spray == 'D')$count)
[1] 4.916667
> mean(subset(InsectSprays,spray == 'E')$count)
[1] 3.5
> mean(subset(InsectSprays,spray == 'F')$count)
[1] 16.66667
```