Question 5.24

The relationship between age-adjusted death rate and animal fat intake is positive, linear, absent of outliers, and very strong (r=0.949).

Question 5.25

The relationship between total dissolved solids and discharge in Fish Creek is mostly positive and mostly linear (Figure 1). Several outliers are apparent of which the three most prominent are the two with a total dissolved solids greater than 800 mg/L and discharge less than 600 cfs and the one point with a total dissloved solids of 0 mg/L and a discharge at approximately 250 cfs (Figure 1). The strength, excluding the outliers, is moderately strong. I did not compute a correlation coefficient because of the presence of the outliers.

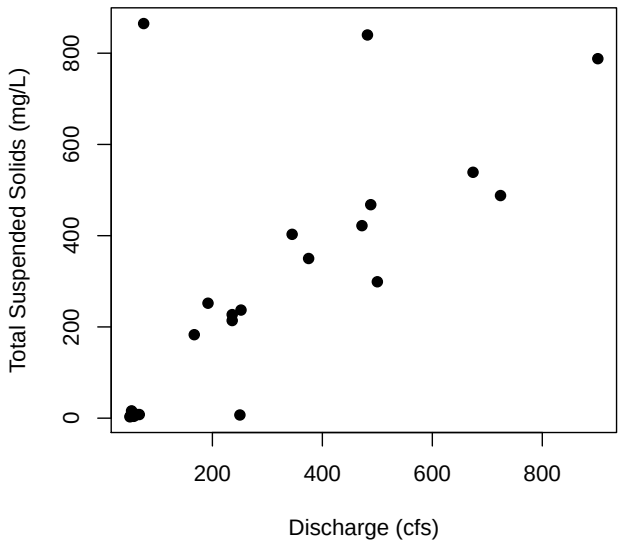

Figure 1. Scatterplot of total dissolved solids versus discharge in Fish Creek.

Question 5.26

a. The frequency table is shown in Table 1.

Table 1. Frequency table for the fire-blight data. Note that A=no action (control), B=removal of the affected branches, C=spraying of foliage with an antibiotic and the removal of affected branches, 1=the tree died in the same year that the disease was noticed, 2=the tree died after 2-4 years, 3=the tree died after 4 years.

	1	2	3
A	5	2	0
B	3	3	2
C	2	4	3

b. The row percentage table is shown in Table 2.

Table 2. Row percentages table for the fire-blight data. Abbreviations are given in Table 1.

	1	2	3	Sum
A	71.4	28.6	0.0	100.0
B	37.5	37.5	25.0	100.0
C	22.2	44.4	33.3	100.0

c. The column percentage table is shown in Table 3.

Table 3. Column percentages table for the fire-blight data. Abbreviations are given in Table 1.

	1	2	3
A	50.0	22.2	0.0
B	30.0	33.3	40.0
C	20.0	44.4	60.0
Sum	100.0	100.0	100.0

d. The table percentage table is shown in Table 4.

Table 4. Table percentages table for the fire-blight data. Abbreviations are given in Table 1.

	1	2	3	Sum
A	20.8	8.3	0.0	29.2
B	12.5	12.5	8.3	33.3
C	8.3	16.7	12.5	37.5
Sum	41.7	37.5	20.8	100.0

e. The percentage of all trees in Treatment A that were dead within the first year (i.e., outcome=1) is 71.4% (Table 2).

f. The percentage of ALL trees that were in Treatment A AND were dead within the first year (i.e., outcome=1) is 20.8% (Table 4).

g. The percentage of the trees in the control treatment (treatment A) that died after four years (i.e., outcome=3) is 0.0% (Table 2).

h. The percentage of the trees that died after 2-4 years (outcome 2) that were in the control treatment (treatment A) is 22.2% (Table 3).

i. The percentage of all trees that were dead within the first year is 41.7% (Table 4).

Appendix – R Commands

```
> d <- read.table("http://www.ncfaculty.net/dogle/R/Data_Master/R/FishCrNWaterQuality.txt",
  header=TRUE)
> plot(SuspSed~DschrgCFS,data=d,pch=19,xlab="Discharge (cfs)",
  ylab="Total Suspended Solids (mg/L)")
```

```
> treat <- c("A","A","A","A","A","B","B","B","C","C","A","A","B","B","B",
  "C","C","C","C","B","B","C","C","C")
> out <- factor(c(1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3))
> d2 <- data.frame(treat,out)
> tbl <- table(d2$treat,d2$out)
> row.tbl <- percTable(tbl,margin=1)
> addMargins(row.tbl,margin=1)
> col.tbl <- percTable(tbl,margin=2)
> addMargins(col.tbl,margin=1)
> perc.tbl <- percTable(tbl)
> addMargins(perc.tbl)
```

Notes From Professor

- Note that computing the correlation in 5.25 is inappropriate because of the outliers in the data. However, if you were to compute the correlation you would have to use the code below, including the `use="pairwise.complete.obs"` argument because there are missing data in the *SuspSed* variable. **Again, this is inappropriate in this situation**.

  ```
  > cor(d$SuspSed,d$DschrgCFS,use="pairwise.complete.obs")
  [1] 0.720706
  ```

- In 5.25, you must explicitly state where the outliers are located. In this case it is adequate to note that they are "in the upper-left" corner of the plot. Alternatively, you could note the approximate coordinates of the points.
- In 5.25, it is correct to not calculate or report the correlations coefficient because of the presence of outliers. However, you still need to comment on the strength of the relationship. Your comment will be more subjective based on your interpretation of the clustering of the points but it still needs to be made.
- In 5.26 it probably would have been easier to enter the data into Excel, save it as a text file, and then use `read.table()` to load it into R.
- Especially in 5.26, sentences cannot begin with a number (e.g., you cannot say "22.2% of ...."). You must reorganize the sentence so that it does not begin with a number.
- Make sure to label and refer to tables.