1. (a)

	B	B^c	
A	0.1	0.3	0.4
A^c	0.4	0.2	0.6
	0.5	0.5	1

	C	C^c	
B	0.15	0.35	0.5
B^c	0.45	0.05	0.5
	0.6	0.4	1

	C	C^c	
A	0.15	0.25	0.4
A^c	0.45	0.15	0.6
	0.6	0.4	1

(b)
```
install.packages("VennDiagram")
library("VennDiagram")
draw.triple.venn(.4,.3,.25,.15,.1,.12,.04,category=c("A","B","C"),
fill=c("Yellow","Blue","Red"))
```
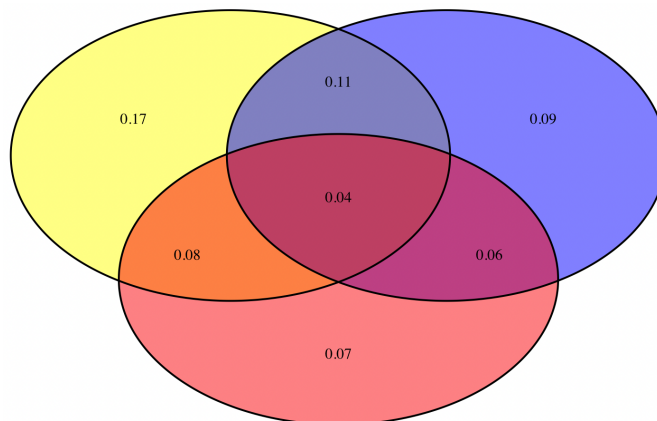

2. (a) Define the events A: you have smallpox; B: you have spots. The information given is

$$P(A) = 0.001; \qquad P(B|A') = 0.081; \qquad P(B|A) = 0.9.$$

What you want to know is, given that you have the spots, what is the probability that you have smallpox? In the notation we have adopted above, that is $P(A|B)$. Apply Bayes' Rule:

$$\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} \\
&= \frac{0.9(0.001)}{0.9(0.001) + 0.081(1 - 0.001)} \\
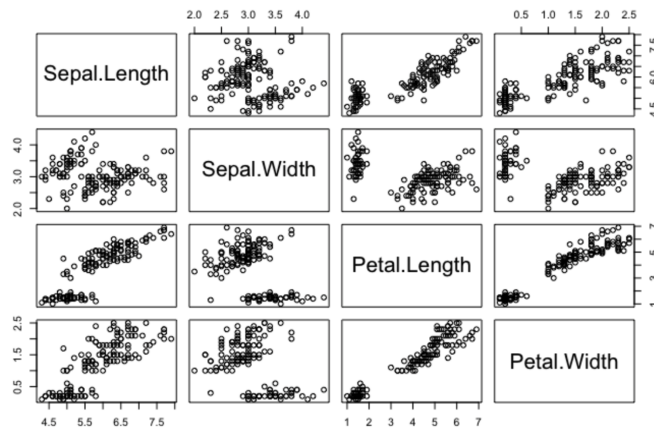&\approx 0.011
\end{aligned}$$

So there is only a 1.1% chance that you have smallpox. You should find a different family doctor.

(b) Define the events A: the patient has chickenpox and B: the patient has spots. By the same method as (a), we find

$$\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} \\
&= \frac{0.8(0.1)}{0.8(0.1) + 0.081(0.9)} \\
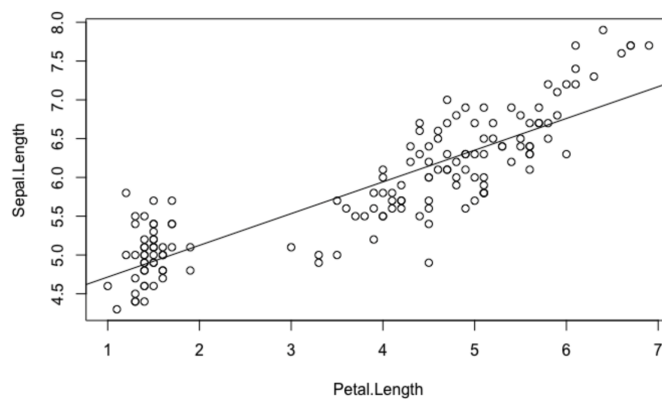&\approx 0.52
\end{aligned}$$

You are about 52% confident that the patient has chickenpox.

3. (a) `pairs(iris[,1:4])`

(b) For instance if you choose sepal length vs. petal length:

```
plot(Sepal.Length~Petal.Length,data=iris)
abline(lm(Sepal.Length~Petal.Length,data=iris))
```


There are no outliers nor influential points. An outlier would be far away from the line and the line would still be well-suited to all the rest of the points. An influential point would cause the line to be sloped significantly differently than it would be if the point were not there.

(c) `cor(iris[,1],iris[,3])`
[1] 0.8717538

The correlation is quite strong and positive, since the coefficient is close to 1. There is a linear relationship.

(d) `t<-lm(Sepal.Length~Petal.Length, data=iris)`
`summary(t)`

```
Call:
lm(formula = Sepal.Length~Petal.Length, data = iris)


Residuals:
Min 1Q Median 3Q Max
-1.24675 -0.29657 -0.01515 0.27676 1.00269


Coefficients:
Estimate Std.  Error t value Pr(>|t|)
```

```
(Intercept) 4.30660  0.07839 54.94 <2e-16 ***
Petal.Length 0.40892  0.01889 21.65 <2e-16 ***
---
Signif.  codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.'  0.1 ' ' 1


Residual standard error:  0.4071 on 148 degrees of freedom
Multiple R-squared:  0.76, Adjusted R-squared:  0.7583
F-statistic:  468.6 on 1 and 148 DF, p-value:  < 2.2e-16
```

The y-intercept is 4.3066 and the slope is 0.40892, so the equation of the best-fit line is

$$y = 0.40892x + 4.3066.$$

(e) For instance a petal length of 8 is reasonable, so the corresponding sepal length should be about $y = 0.40892 \cdot 8 + 4.3066 = 7.57796$. Notice, though, that with this formula the petal length will soon overtake the sepal length, which might not be reasonable for this species. It could be that this model is not useful much beyond what is in the plot.