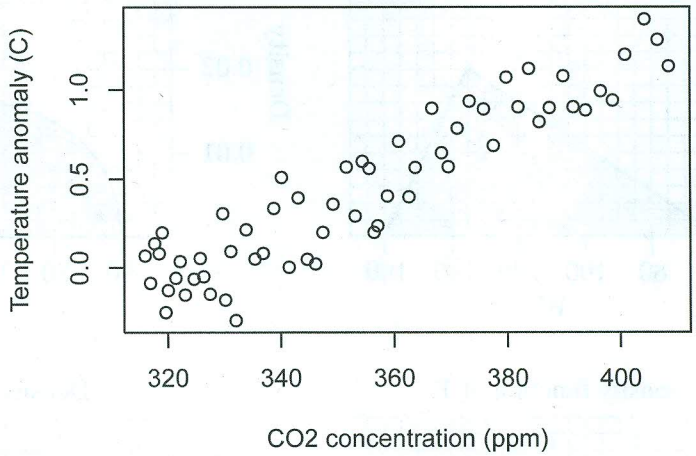
Q2. (10 points) A study was done to investigate the relationship between CO_2 and average temperature. The Mauna Lau observatory has continuously measured the concentration of CO_2 over the last hundred years. Many other observatories have continuously measured temperature. Below we plot the two variables, where temperature is represented as degrees Celsius above expected (temperature anomaly).



(a) What kind of study was this (observational or experimental)?

Observational

(b) What is the implied explanatory variable?

CO_2 concentration

(c) What is the implied response variable?

Temperature anomaly

(d) What association is there between the two variables (positive, negative, or none)?

positive

(e) Based on this study, should we conclude there is a causal relationship between the variables?

No.

(f) Suggest another possible hypothesis than "more CO_2 causes higher temperature anomalies". For example, provide a possible confounding variable.

Possible Confounding variable: time (both just happen to be increasing)
Ozone hole, volcanoes ...
→ Maybe higher temp causes more CO_2.
(both of these are largely discredited)

Q3. (5 points) Complete the contingency table below by assuming A and B are **independent** events.

	A	A^c	total
B	0.1	0.15	0.25
B^c	0.3	0.45	0.75
total	0.4	0.6	1

Independence $\Rightarrow$ $P(B) = P(B|A)$

$= \frac{0.1}{0.4} = 0.25$

Q4. (5 points) A random sample of the bikes on Craiglist (near Boston in February) provided the following prices (in USD):

145 175 240 160 175 222 75 500 299 1300

Make a box plot summarizing these data.

75 145 160 175 175 | 222 240 299 500 1300

Q_1 = 160, median = 198.5, Q_3 = 299

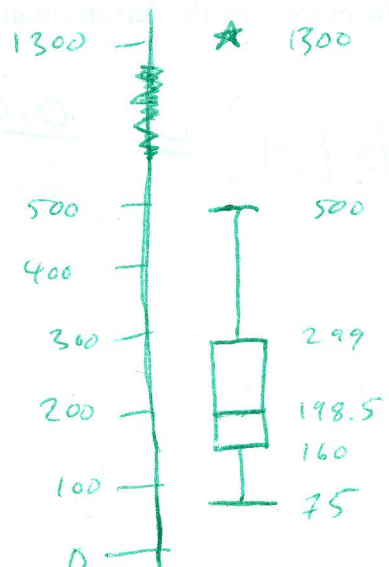
IQR = 299 − 160 = 139

Check for outliers

160 − 1.5 × 139 = −48.5

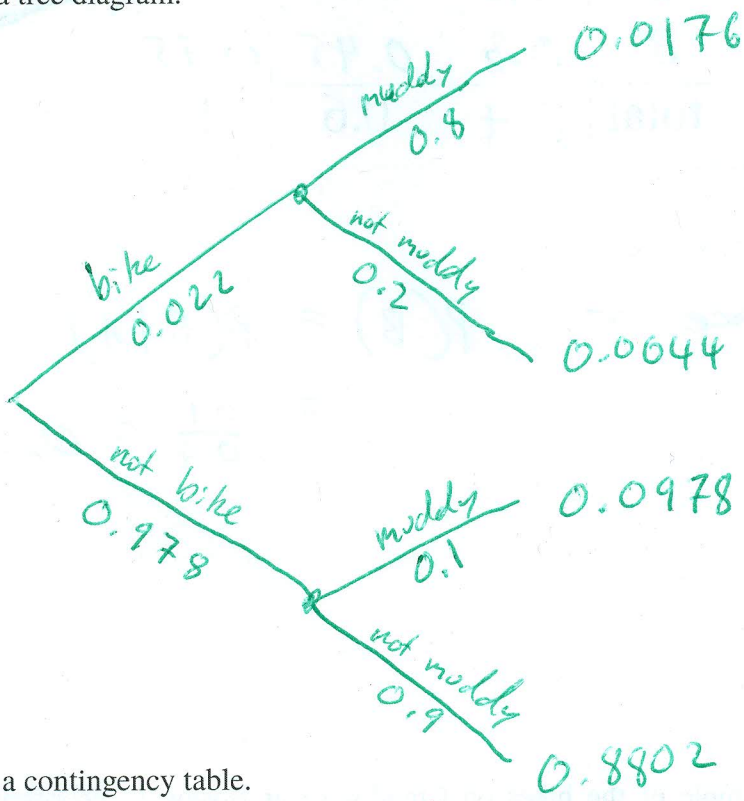
299 + 1.5 × 139 = 507.5

1300 is an outlier!



Q5. (10 points) About 2.2% of Boston commuters use bicycles. If a Boston commuter uses a bicycle, there is an 80% chance their jacket is muddy. If a Boston commuter uses a nonbicycle, there is a 10% chance their jacket is muddy. You see a Boston commuter with a muddy jacket and wonder if they commute via bicycle.

(a) Draw a tree diagram.



(b) Make a contingency table.

	bike	not bike	total
M	0.0176	0.0978	0.1154
M^c	0.0044	0.8802	0.8846
total	0.022	0.978	1

(c) Determine the probability the person commutes via bicycle given their jacket is muddy.

$$P(B|M) = \frac{0.0176}{0.1154} \approx \boxed{0.153}$$

Q6. (10 points) An urn contains marbles. Each marble has a color and a pattern. The frequencies are shown in the contingency table.

	red	green	blue	total
dotted	18	24	15	57
striped	32	16	23	71
checkered	27	19	30	76
filled	15	22	16	53
total	92	81	84	257

(a) What is the probability that a random marble is red?

$\frac{92}{257} \approx 0.358$

(b) What is the probability that a random marble is checkered?

$\frac{76}{257} \approx 0.296$

(c) What is the probability that a random marble is blue and striped?

$\frac{23}{257} \approx 0.089$

(d) What is the probability that a random marble is blue or striped?

$\frac{32+16+23+15+30+16}{257} = \frac{84+71-23}{257} = \frac{132}{257} \approx 0.514$

(e) What is the probability that a random marble is striped given it is blue?

$\frac{23}{84} \approx 0.274$

(f) What is the probability that a random marble is blue given it is striped?

$\frac{23}{71} \approx 0.324$