Substitution Ciphers
A monoalphabetic substitution (also called a simple substitution) is just a permutation of the alphabet.

Ex 1. The ATBASH cipher is a simple substitution:

Table 1: ATBASH

plain	A	B	C	D	E	F	G	H	I	J	K	L	M
cipher	Z	Y	X	W	V	U	T	S	R	Q	P	O	N
plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cipher	M	L	K	J	I	H	G	F	E	D	C	B	A

Ex 2. Simple substitution with key word: If sender and receiver agree on a keyword then a simple substitution can be generated from that keyword. Consider the key word DEER-HOOF. First write the word without letter repetitions DERHOF. Place this at the beginning of the table and then proceed alphabetically to fill in the rest.

Table 2: keyword DERHOF

plain	A	B	C	D	E	F	G	H	I	J	K	L	M
cipher	D	E	R	H	O	F	A	B	C	G	I	J	K
plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cipher	L	M	N	P	Q	S	T	U	V	W	X	Y	Z

NOTE: Simple substitutions will always have inverses since permutations are one-to-one. This is the reason we cannot use repeated letters in a keyword, or else the cipher will not be invertible.

Ex 2. One can also use a key word to generate a simple substitution via a columnar transposition similar to the one we saw in the ADFGVX cipher. We use the same keyword DERHOF. We write the entries of table 2 as follows:

$$\begin{pmatrix} D & E & R & H & O & F \\ A & B & C & G & I & J \\ K & L & M & N & P & Q \\ S & T & U & V & W & X \\ Y & Z & & & & \end{pmatrix}$$

Now one can write a table for the substitution using some predetermined rule. For instance one can just write it column by column or one can write it alphaabetically column by column. In the book, example 2.3.2 demonstrates the latter so we will do an alphabetic columnar transposition. So we start with the column which starts with D, and then E, then F, ...

Table 3: keyword DERHOF alphabetic columnar transposition

plain	A	B	C	D	E	F	G	H	I	J	K	L	M
cipher	D	A	K	S	Y	E	B	L	T	Z	F	J	Q
plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cipher	X	H	G	N	V	O	I	P	W	R	C	U	M

Using this table we can encipher SPACE IS THE PLACE as:

OGDKY TO ILY GJDKY

NOTE: Table 3 seems to be a better substitution compared to table 2 since there are no repeats.

Cryptanalysis of a Substitution Cipher

Since simple substitutions just replace one letter for another, a letter frequency anaysis will give information. Read example 2.3.3 in Barr for a thorough example of how to use letter, diagraph and trigraph frequencies for cryptanalysis. We will work carefully through question 4 in section 2.3.

4. We have the following cipher text enciphered using a columnar keyword transposition:

ZRXEKGREU LJP KUOUJ TULEK LIR RXE ZLSBUEK AERXIBS ZRESB RJ SBHK GRJSHJUJS L JUF JLSHRJ GRJGUHOUP HJ NHAUEST LJP PUPHGLSUP SR SBU WERWRKHSHRJ SBLS LNN DUJ LEU GEULSUP UCXLN

Table 4: cipher letter count

cipher	U	S	J	R	L	E	H	P	B	G	K	X	N
count	18	15	14	14	13	11	9	7	6	6	5	5	4
cipher	Z	A	I	O	T	W	C	D	F	M	Q	V	Y
count	3	2	2	2	2	2	1	1	1	0	0	0	0

Throughout this example we will use lowercase for plain text and capitals for cipher text. We also use tables 2.6, 2.7 and 2.8 from Barr in our analysis. Using this letter count our

first guess is that e → U and t → S. Under this substitution the cipher text SBU becomes tBe. Since this is the only trigraph which starts with t and ends with e we may assume h → B.

ZRXEKGREe LJP KeOeJ TeLEK LIR RXE ZLtheEK AERXIht ZREth RJ thHK GRJtHJeJt L JeF JLtHRJ GRJGeHOeP HJ NHAeEtT LJP PePHGLteP tR the WERWRKHtHRJ thLt LNN DeJ LEe GEeLteP eCXLN

The cipher text L must correspond to the plain text a or i. This along with the apperance of thLt means that we must have a → L. Under this substitution the cipher text LNN becomes aNN. So we may also assume l → N

ZRXEKGREe aJP KeOeJ TeaEK aIR RXE ZatheEK AERXIht ZREth RJ thHK GRJtHJeJt a JeF JatHRJ GRJGeHOeP HJ lHAeEtT aJP PePHGateP tR the WERWRKHtHRJ that all DeJ aEe GEeateP eCXal

Since "and" is the second most common trigraph (and the only remaining trigraph remaining in table 2.8 which starts with a) we assume: and → aJP...i.e. n → J and d → P.

ZRXEKGREe and KeOen TeaEK aIR RXE ZatheEK AERXIht ZREth Rn thHK GRntHnent a neF natHRn GRnGeHOed Hn lHAeEtT and dedHGated tR the WERWRKHtHRn that all Den aEe GEeated eCXal

Lets list what we have so far:

plain	a	b	c	d	e	f	g	h	i	j	k	l	m
cipher	L			P	U			B				N	
plain	n	o	p	q	r	s	t	u	v	w	x	y	z
cipher	J						S						

Recall that the most frequent cipher letters and plain text letters are

cipher: U S J R L E H P B G K X N
plain : e t a o i n s h r l d c u

Throwing out the letters we have already used we have left

cipher: R E H G K X
plain : o i s r c m u.

The digraphs Rn, Hn and tR imply that H and R must correspond to vowels. The only way this is possible is via: i → H and o → R.

ZoXEKGoEe and KeOen TeaEK aIo oXE ZatheEK AEoXIht ZoEth on thiK Gontinent a neF nation GonGeiOed in liAeEtT and dediGated to the WEoWoKition that all Den aEe GEeated eCXal

Although the plaintext is probably clear to everyone we continue the analysis. The text "aIo" seems problematic...but trying all possible remaining substitutions (note that d is already taken) for I gives only the plain text g. Thus we have g → I.

ZoXEKGoEe and KeOen TeaEK ago oXE ZatheEK AEoXght ZoEth on thiK Gontinent a neF nation GonGeiOed in liAeEtT and dediGated to the WEoWoKition that all Den aEe GEeated eCXal

The most frequent remaining letters are

cipher: E G K X
plain : s r c m u.

Assuming one of the plain text letters "s r l c m u" is enciphered as E, the block aEe shows that r → E. Now we assume s enciphers as G, K or X. s → G does not make sense with the text "Greated". s → K does not seem to give any contradictions. If s → X then we have a contradiction with the text "oXr". So we assume s → K. This gives

ZoXrsGore and seOen Tears ago oXr Zathers AroXght Zorth on this Gontinent a neF nation GonGeiOed in liAertT and dediGated to the WroWosition that all Den are Greated eCXal

We now just write the plaintext

fourssore and seven years ago our fathers brought forth on this continent a new nation conceived in liberty and dedicated to the proposition that all men are created equal.

ZRXEKGREU LJP KUOUJ TULEK LIR RXE ZLSBUEK AERXIBS ZRESB RJ SBHK GRJSHJUJS L JUF JLSHRJ GRJGUHOUP HJ NHAUEST LJP PUPHGLSUP SR SBU WERWRKHSHRJ SBLS LNN DUJ LEU GEULSUP UCXLN.

Thus, this was enciphered using the following substitution

plain	a	b	c	d	e	f	g	h	i	j	k	l	m
cipher	L	A	G	P	U	Z	I	B	H			N	D
plain	n	o	p	q	r	s	t	u	v	w	x	y	z
cipher	J	R	W	C	E	K	S	X	O	F		T	

The four missing letters are M, Q, V, Y. Now we wish to find the key. Assume the key is of length 3. Then the first two columns are of length 9 and the last of length 8. This would look like

$$\begin{pmatrix} L & ? & K \\ A & ? & S \\ G & N & X \\ P & D & O \\ U & J & F \\ Z & R & ? \\ I & W & T \\ B & C & ? \\ H & E & . \end{pmatrix}$$

since M, Q, V, Y don't fit into the middle position L ? K we know that the key word is not of length 3. If the key is of length 4, then the first two columns are of length 7 and the latter two are of length 6 ($7+7+6+6=26$). This yields the keyword LBRX which is not English. A key word of length 5 will give the first column length 6 and the remaining 4 of length 5. This gives the keyword LINCO:

$$\begin{pmatrix} L & I & N & C & O \\ A & B & D & E & F \\ G & H & J & K & ? \\ P & ? & R & S & T \\ U & ? & W & X & ? \\ Z & & & & \end{pmatrix}$$

Thus, it seems that the keyword was LINCOLN.

Outline for 2.3, question 5

I suggest using a text program for this question. Type the cipher text in CAPITALS and whenever you make a substitution make a copy of the text first and then substitute into the copied version. This way, if you need to change any substitution that you made you can go back to the original. Use capital letters for cipher text and lower case for plain text.
Hints: $c \to L$, $s \to Q$ and $m \to R$.

1. Calculate letter frequency. There is a website for this

http://www.10ticks.co.uk/s_codebreaker_letter.asp

2. To identify the word *liberty* assign cypher text to the plaintext letters e and t (in the most obvious way using letter frequency).

3. Make the replacements as dictated by the identification of the work *liberty*. (it helps to keep track of the remaining plain text and cipher text letters. Also, keep track of your list of substitutions.

4. The plain text diagraph *er* shold appear...extend this by one letter using the table of the most frequent trigraphs (in particular...what is the most frequent trigraph containing *er*?)

5. Figure out how the next two most frequent plain text letters encipher as the remaining most frequent cipher text letters.

6. Try to detect possible word breaks.

7. The rest is guess work using the remaining letters. Good Luck!