Q1: Assuming you have R installed (if not install it). Load up the various packages you need for using the wordcloud packages:

a. Carry out the commands shown in the practical notes:

```
> library(wordcloud)
> library(tm)
> wordcloud("May our children and our childrens children to a thousand generations
 continue to enjoy the benefits conferred upon us by a united country and have cau
se yet to rejoice under those glorious institutions bequeathed us by washington an
d his compeers.",colors = brewer.pal(6,"Dark2"),random.order = FALSE)
```

Plot:

After carrying out the commands ,there seems to be some problems about the outcomes and some warnings emerge in the console. Such as “rejoice could not be fit on page. It will not be plotted.” So I tried to plot with a larger device using :

dev.new(width = 1200, height = 1200, unit = "px")

Now the plot turns to be normal:

b. When you have done this, report the list of the words from the original quote that are included in the wordcloud and the list of those that are not. Report why do you think some are excluded and others included?

Words that are included in the wordcloud:

children, generations, thousand, enjoy, compeers, Washington, institutions, cause, united,

bequeathed, upon, yet, benefits, childrens, continue, country, rejoice, conferred, may, glorious.

Words that are not included in the wordcloud:

our, and, to, a, the, us, by, have, under, his.

I think the wordcloud package will filter out the stopwords automatically such as and, to, a, that will not make too much influence to the outcomes.

c. Now, check your theory about what the wordcloud package included and excluded. Put in your own word-list together (30-50 words) and check what wordcloud includes and excludes? Report whether your initial theory was right or wrong and why?

My own word-list: "This essay aims to demonstrate the developments of surgical robots, identify the benefits and limitations of each technology, and through analyzing the situation over decades to reach a conclusion about what are the challenges and risks will be in the development of surgical robots and feasible measures to meet the challenges with."

As we can see, that the words like "to", "with", "of", "and", "the" has not been included, because they are too often and not important.

d. Again, using your word-list add more repeated words and see what happens? Can you change the package's to make it more inclusive of the words in the word-list?

My word-List: “This essay aims demonstrate to demonstrate the developments of surgical robots, identify the benefits and limitations of limitations each technology, and through analyzing the situation over decades to reach a conclusion about what are the challenges and risks will be in the development of surgical robots and feasible measures.”

After increase the frequency of demonstrations and limitations, the center of the plot has been changed to four words.

Q2: Find the Google Ngram Viewer online and do the following with it:

a. Put in “Mark Keane” as a search term and explain the peaks that appear in the graph over time.

This graph illustrates the frequency of “Mark Keane” in the book between 1940 to 2008. There are two peaks in the chart. The first peak appears in 1973 which is 0.0000002262% and the second one appears around 2000 which is 0.0000001544%.

b. Put your own name in and describe what happens, explaining where the hits are coming from.

This graph illustrates the frequency of “Bingbing” in the book between 1800 to 2008. The appearance of the first peak is because this name appears in “Indonesian journal for natural science” as a writer and some other journal. The increasement of the frequency of this name is because that it is a common Chinese name and there are two famous stars in China named Bingbing, so it increased dramatically after 2000.

c. Pick a word that you think is a recent introduction into the English language (like “exit strategy”) and plot its emergence, showing the graphs. If it actually emerges before you thought, explain why?

This graph illustrates the frequency of “surgical robots” which is a new word in the book between 1900 to 2008. It actually emerges from around 1965. The reason why this new word appeared in that years is because it was described in the magazine “Galaxy Magazine: Science Fiction”.

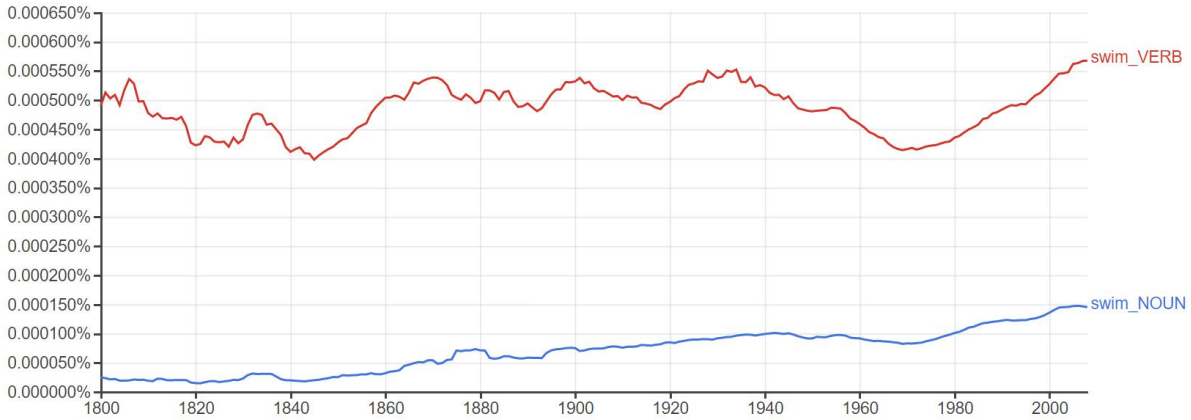
d. Describe some of the effects of smoothening these graphs with different values?

The greater the smoothing extent is, the more smooth and general the line will become, moreover the data will be averaged more. On the other hand, if it is set to 0, the data will be raw and relatively accurate.

e. Do a comparison between 3 or more related terms to see how their relative frequencies have changed over time*. Is there anything surprising about how these terms differ in their frequency and, if so, why? Why do you think the frequencies vary in the way they do.

This graph illustrates the frequency of three bands: “Led Zeppelin”, “Pink Floyd”, “Nirvana” in the year between 1900 to 2008. It appears that Led Zeppelin and Pink Floyd have the similar trends. While nirvana not only a famous band ,but also a religious vocabulary, therefore it always has the most frequent appearance.

f. Use the syntactic tags in a search for two words that are the same but syntactically different (e.g., fish-verb, fish-noun; do not use fish) and report what you find.



This graph illustrates the frequencies of “swim” as a noun and as a verb in the book between 1800 to 2008. It shows that the appearance of swim as a noun has been increasing slightly from 1800. While it as a verb has always been used more than swim as a noun and it has been fluctuating in this period.

g. **Think of some major cultural change that has happened over the last 500 years and some words that could denote to ths event/events. Check these words of the relevant time period. Report what you find.**

This graph illustrates the frequencies of “minimally invasive surgery”, “surgical robots”.It shows that they appeared almost at the same time. The background is that the appearance of the surgical robots facilitate the way the surgeons perform minimally invasive surgery. And this plot shows that the minimally invasive surgery has more frequency than surgical robots.

Q3: Using an Excel spreadsheet set up your own list of 15 words and give each a made-up frequency between 0 and 2000 for each of three years (2010, 2011, 2012). Now perform two different normalisations on them:

a. Method1: produce a normalised frequency for each word in each year, using the total N of words in all years.

b. Method2: produce a normalised frequency for each word in each year, using the N of words in a given year

c. Does normalising by method1 or method2 make a big difference to the scores produced? Graph the difference and comment on it.

Name	2010	2011	2012	Total
Joy	1724	429	316	2469
Sad	380	722	1878	2980
Glad	1766	1799	755	4320
Wealthy	1129	1575	1870	4574
Poor	1084	1013	1084	3181
Greed	1558	238	1662	3458
Happy	334	1305	521	2160
High	214	491	1085	1790
Low	1509	546	390	2445
Long	1672	1820	23	3515
Short	780	1300	67	2147
Tall	830	677	540	2047
Quick	970	207	1270	2447
Slow	1336	1356	505	3197
Fast	1510	923	1969	4402
Total	16796	14401	13935	45132

Method1	2010	2011	2012
Joy	0.698258404	0.173754557	0.127987039
Sad	0.127516779	0.242281879	0.630201342
Glad	0.408796296	0.416435185	0.174768519
Wealthy	0.246829908	0.34433756	0.408832532
Poor	0.340773342	0.318453317	0.340773342
Greed	0.450549451	0.068825911	0.480624639
Happy	0.15462963	0.604166667	0.241203704
High	0.119553073	0.274301676	0.606145251
Low	0.617177914	0.223312883	0.159509202
Long	0.475675676	0.517780939	0.006543385
Short	0.363297625	0.605496041	0.031206334
Tall	0.405471422	0.330727894	0.263800684
Quick	0.39640376	0.08459338	0.519002861
Slow	0.417891774	0.424147638	0.157960588
Fast	0.343025897	0.209677419	0.447296683
Total	0.372152796	0.319086236	0.308760968

Method	2010	2011	2012
Joy	1.876268058	0.544537924	0.414518196
Sad	0.3426463	0.759299061	2.041065445
Glad	1.098463589	1.305086645	0.566031775
Wealthy	0.663248834	1.079136363	1.324106912
Poor	0.915681261	0.998016463	1.103680119
Greed	0.415500384	1.89342754	0.78119882
Happy	0.415500384	1.89342754	0.78119882
High	0.321247278	0.859647472	1.963153749
Low	1.658399239	0.699851195	0.516610644
Long	1.278173053	1.622699072	0.021192399
Short	0.976205549	1.897593731	0.101069558
Tall	1.089529424	1.036484365	0.85438482
Quick	1.065163996	0.26511134	1.680921213
Slow	1.122903758	1.329257081	0.511595067
Fast	0.921734032	0.657118345	1.448682735
Total	1	1	1

It seems that the results of Method 1 and Method 2 after normalization have not changed much.