Edit distance

• String alignment. Gaps can be put between characters
• There is a cost associated with a gap \(+2\)
• There is a cost associated with two character being different \(+1\)

S U N D A Y _ _ cost  = 9
S A T U R D A Y

We want to find the optimal alignment
Can we try all possible alignments?

NO
Way too many possibilities
(makes for an interesting CIS160 question)
Recursion to the rescue!

To solve this consider what the possibilities are for the first characters in the optimal alignment

F or - or F
-
F or F
Recursion to the rescue!

F or - or F
- F F

In case 1 – optimally align RIEND and FIEND and remember to add 2 for blank

Case 2 – optimally align FRIEND and IEND and remember to add 2 for the blank

Case 3 – optimally align RIEND and IEND (the two Fs match so no additional penalty)
Recursive formula

\[
\text{Editdist}(\text{string1}, \text{string2}) = \\
\text{minimum of} \\
\text{editdist}(\text{rest of string 1}, \text{string2}) + 2 \\
\text{editdist}(\text{string1}, \text{rest of string 2}) + 2 \\
\text{editdist}(\text{rest of string 1}, \text{rest of string 2}) + \text{either} \\
0 \text{ or } 1 \text{ depending on whether the first character of} \\
\text{string 1 and string 2 are the same.}
\]
Recursion is slow for this problem

We are recomputing the same thing over and over and over again
Alternative approach

Using a look up table