Greedy Algorithms

① Huffman Coding

② Brute Force - Greedy Blend

ⓐ Walls

ⓑ Welcome Party

③ Code Huffman Coding

File Compression

Imagine input is using fixed-length coding (ascii codes - each 8 bits)

What about variable length coding?

A - 11
B - 101
etc.

What's good?

Use fewer bits for more frequent chars.

What's bad?

When reading, we don't know when a char ends!!!

Can't figure out if 11 is A or the start of B...

PROBLEM - A - 11
C - 01 B - 1101

Huffman Coding

Input - fixed encoding set of chars

Output - Variable length encoding such that the prefix for one code isn't another code.

(A - 11 → B's prefix 11
B - 1101 was the code for A)

Huffman ~~minimizes~~ minimizes encoded file size!

Note: each file encoded in a diff way.

① Get a freq chart of all chars.

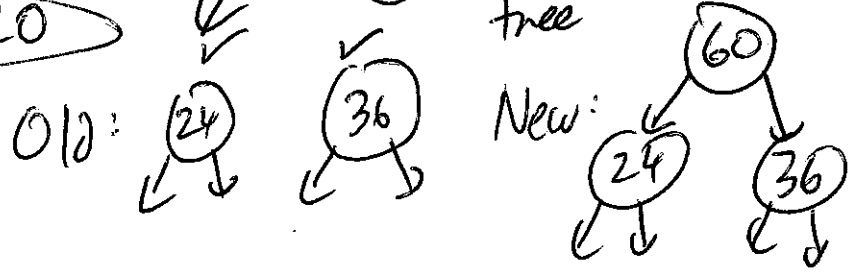
- ✓ A - 5
- ✓ B - 30
- ✓ C - 15
- D - 100
- ✓ E - 10
- F - 65
- G - 45
- ✓ H - 20

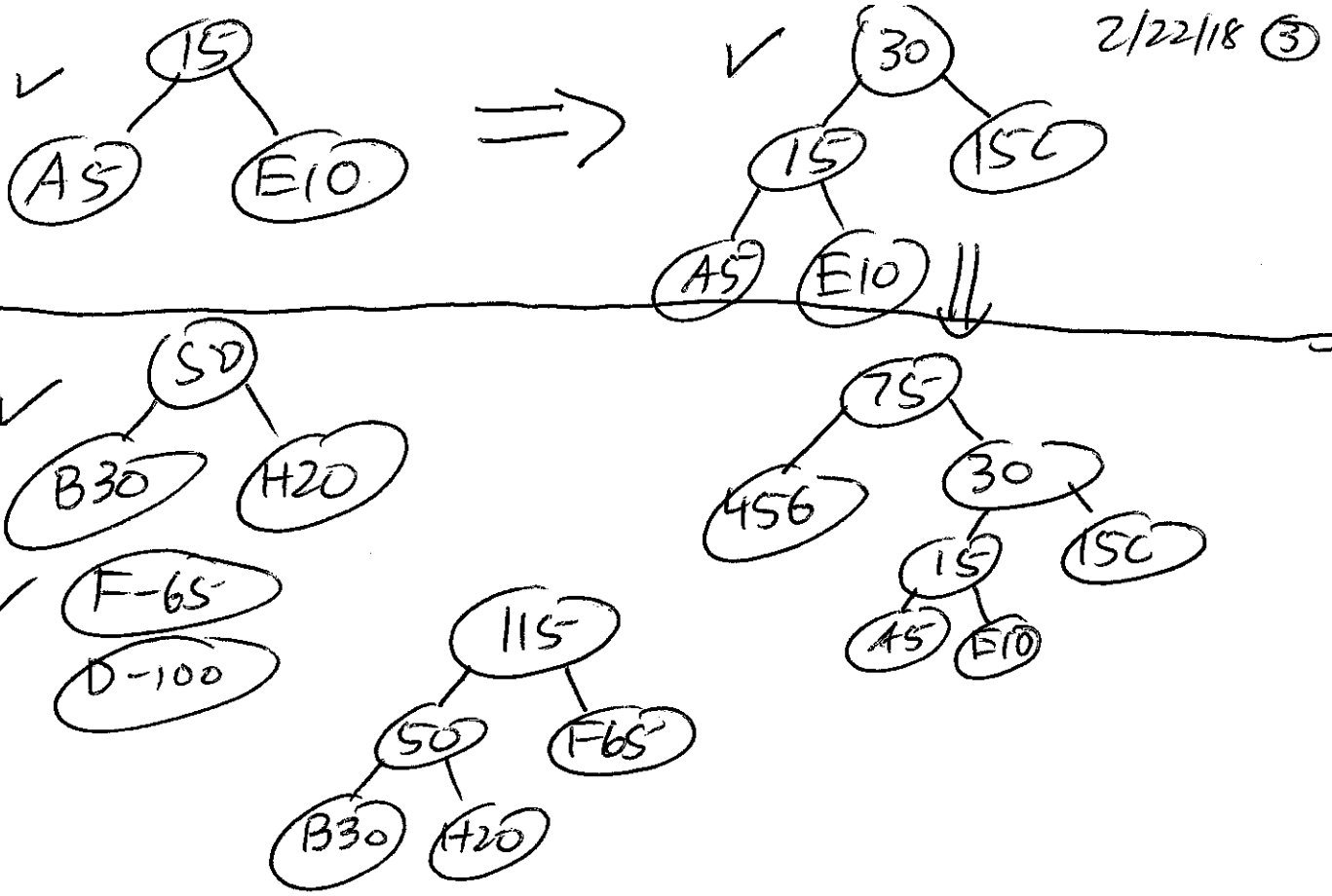
② Make a node for each letter/freq pair

③ Repeat until there is one tree

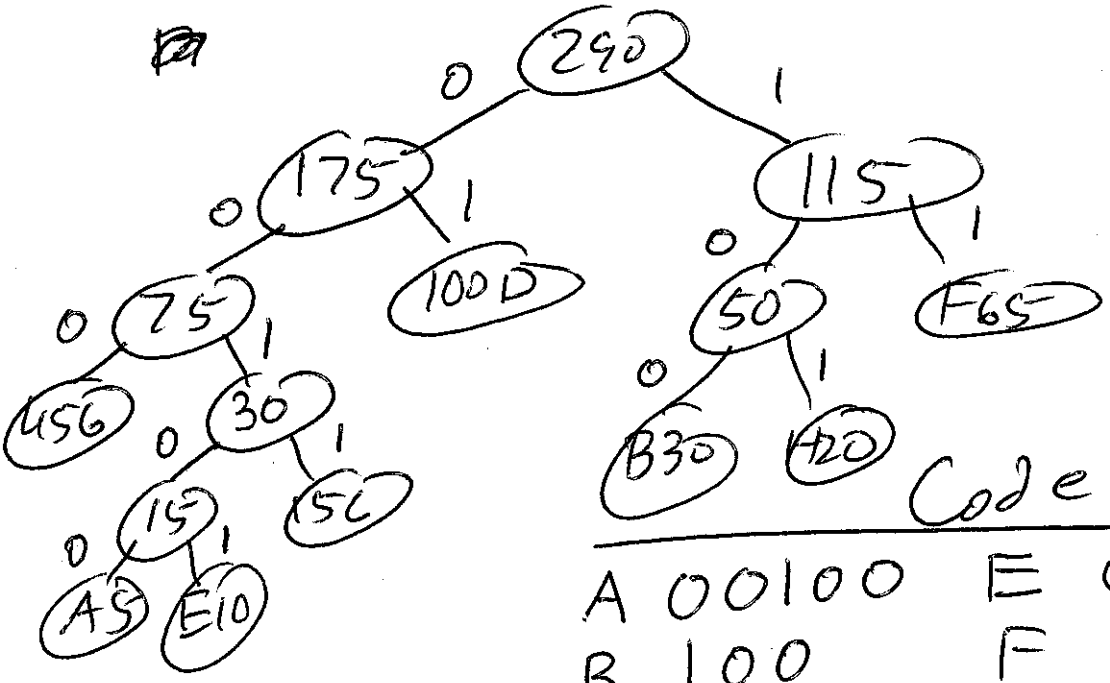
④ find 2 minimum ~~nodes~~ trees (value @ root)

⑤ fuse them into one tree





2 TREES → 1 TREE



Code

A	00100	E	00101
B	100	F	11
C	0011	G	000
D	01	H	101

2/22/18 (4)

BEFORE

min k
$2^k \geq D$ → # unique chars

290 chars × 3 bits/char
= 870 bits

AFTER

	bits		
A-5	25	115	375
B-30	90	260	375
C-15	60	180	375
D-100	200	195	750
E-10	50		
F-65	130		
G-45	135		
H-20	60		

SAVED 120 bits

A +10 D -100 F -65 } -120 SAVING
C +15 E +20

120 bits!

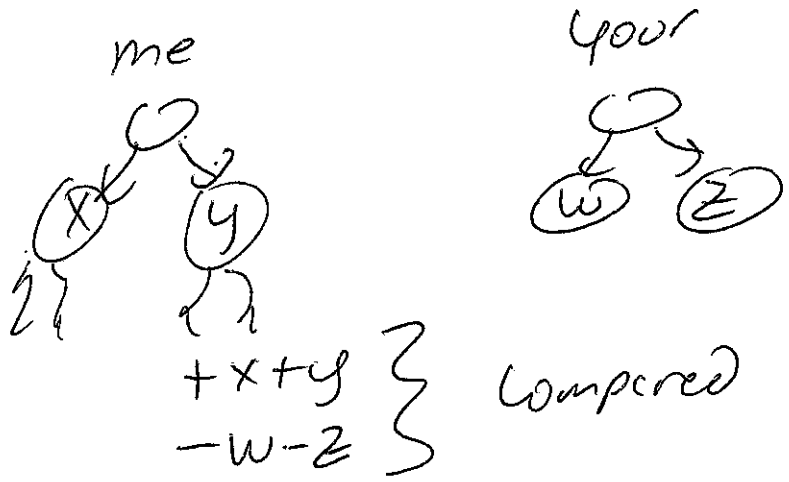
2/22/18 ©

Why is no code a prefix of another code? ALL LETTERS ARE LEAF NODES.

WHY IS THIS OPTIMAL?

---

① When we merse 2 trees, we are adding 1 char to each of those code



2/22/18 (6)

J: Walls

There are a number of research stations on a featureless patch of desert, which can be modeled as a Cartesian plane. Each station is located at some point (*x*,*y*) where *x* and *y* are even integers. For security reasons, sufficiently long and high walls are to be constructed to separate the stations so that no station is visible from any of the other stations. A wall may only be constructed along a North-South or East-West line. A vertical wall may be built at an odd x-coordinate, and a horizontal wall may be built at an odd y-coordinate. Since the stations are located at even valued coordinates, and the walls are built along odd valued coordinates, no wall can ever touch a station. The walls are always long enough to completely separate stations on one side from stations on the other side.

Given a list of stations, you must determine the smallest number of walls that need to be constructed.

Input

There will be several test cases in the input. Each test case will begin with an integer *n* ($2 \le n \le 100$), which is the number of stations. The next *n* lines each will contain two integers *x* and *y* ($0 \le x,y \le 36$), separated by a single space, indicating the (*x*,*y*) location of a station. The *x* and *y* values are guaranteed to be even. Within a test case, all (*x*,*y*) locations will be unique. The last test case will be followed by a line with a single **0**.

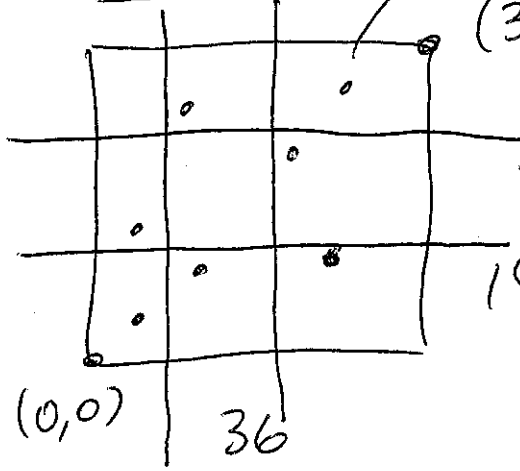
Output

For each test case, output a single integer, indicating the smallest number of walls that can prevent the given *n* stations from seeing each other. That is, a straight line-segment joining any two stations must be intersected by at least one wall. Output no extra spaces, and do not separate answers with blank lines.

2/22/18 (5)

Sample Input	Sample Output
4 12 12 4 8 8 6 2 4 4 0 0 4 4 10 8 14 6 0	2 3

Walls



18 possible

Some animals
(x, y)
x is odd
y is odd

Barriers
Horizontal or Vertical
at even coordinates

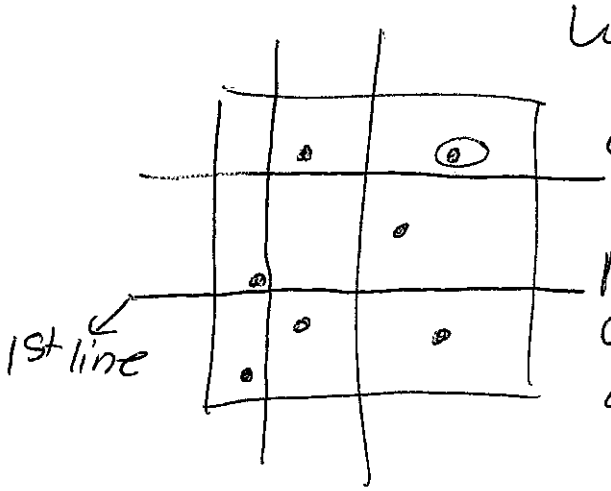
Goal: What is min # of walls to build so no 2 animals can attack each other

First Idea: Brute Force There are 36 possible walls try each subset of walls, Problem - run time $2^{36} \times$ #dots

TOO SLOW

HARD TO ARGUE ANY GREEDY IS CORRECT

New Idea - BRUTE FORCE VERTICAL WALLS (2^{18}) - OK



← one example of a subset of vertical walls

NEW GOAL: Get best answer assuming these vertical walls are fixed.

DRAW MIN HORIZ LINES TO PUT EVERYONE IN THEIR OWN PEN

2/22/18 (9)

→ GO BOTTOM TO TOP, DON'T DRAW A LINE UNTIL YOU ARE FORCED TO.

OVERALL SOLN

TRY ALL COMBOS OF VERTICAL WALLS, FOR EACH DRAW HORIZ ~~GGG~~ GREEDILY.

OF ALL 2^{16} possibilities, CHOOSE THE BEST (# of dots)

2^{16} × GREEDY TIME

Problem A: Welcome Party

Source file: `welcome.{c, cpp, java}`
Input file: `welcome.in`

For many summers, the Agile Crystal Mining company ran an internship program for students. They greatly valued interns' ability to self-organize into teams. So as a get-to-know-you activity during orientation, they asked the interns to form teams such that all members of a given team either have first names beginning with the same letter, or last names beginning with the same letter. To make it interesting, they asked the interns to do this while forming as few teams as possible.

As an example, one year there were six interns: Stephen Cook, Vinton Cerf, Edmund Clarke, Judea Pearl, Shafi Goldwasser, and Silvio Micali. They were able to self-organize into three teams:

- Stephen Cook, Vinton Cerf, and Edmund Clarke (whose last names all begin with C)
- Shafi Goldwasser and Silvio Micali (whose first names begin with S)
- Judea Pearl (not an interesting group, but everyone's first name in this group starts with J)

As a historical note, the company was eventually shut down due to a rather strange (and illegal) hiring practice---they refused to hire any interns whose last names began with the letter S, T, U, V, W, X, Y, or Z. (First names were not subject to such a whim, which was fortunate for our friend Vinton Cerf.)

Input: Each year's group of interns is considered as a separate trial. A trial begins with a line containing a single integer N, such that $1 \le N \le 300$, designating the number of interns that year. Following that are N lines---one for each intern---with a line having a first and last name separated by one space. Names will not have any punctuation, and both the first name and last name will begin with an uppercase letter. In the case of last names, that letter will have an additional constraint that it be in the range from 'A' to 'R' inclusive. The end of the input is designated by a line containing the value 0. There will be at most 20 trials.

Output: For each trial, output a single integer, k, designating the minimum number of teams that were necessary.

Example Input:	Example Output:
6 Stephen Cook Vinton Cerf Edmund Clarke Judea Pearl Shafi Goldwasser Silvio Micali 9 Richard Hamming Marvin Minskey John McCarthy Edsger Dijkstra Donald Knuth Michael Rabin John Backus Robert Floyd Tony Hoare 0	3 6

ACM Mid-Central Programming Competition 2013

WELCOME PARTY

People @ Party (F. I. L. I)

- SC
- SM
- AB
- AM
- BQ
- RS
- BM

First S's: SC, SM

Last m's: AM, BM

First A's: AB

Last S's: RS

First B's: BQ

All last names A-R first A-Z

→ BRUTE FORCE all 2^{17} combos of last name clubs.

SOLVE PROBLEM FOR A FIXED SET OF Last name clubs.

Last M: SM, AM, BM

Last C: SC

- SC
- SM
- AB
- AM
- BQ
- RS
- BM

(3) Count # of unique first initials amongst unused participants