The Majority problem is to find the element that occurs more than half the time in an array, assuming it exists.

Surprisingly, this can be solved in linear time. Here's a linear time algorithm:

```
LinearFindMajority(A)
t := 0
r := 0
c := A[1]
while t ≠ length(A)
    t := t+1
    if A[t] = c then
      r := r+1
    elsif t = 2*r+1 then
      c := A[t]
      r := r+1
    end if
end while
```

c is the current candidate
t is an index into A[1..n]
r is an upper bound on occurrences of c in A[1..t]

Loop Invariant:

(a) $t_i \le 2r_i$
(b) number of occurrences of c_i in A[1..t_i] is $\le r_i$
(c) number of occurrences of any v OTHER than c_i in A[1..t_i] is $\le t_i - r_i$

We'd like you to present the meat of the partial correctness proof, which is that this loop invariant holds. Depending on how long it's taking in tutorial, you can skip similar cases and leave them as exercises.

Proof

Base case is trivial.

Inductive Step:

Assume it's true for i.
Assume at least i+1 iterations.
Prove for i+1.

For each iteration, $t_{i+1} = t_i + 1$.

For the if, t = t_{i+1}.
In the if conditions, r = r_i.

Corresponding to the if conditions, we'll refer to cases:
I: $A[t_{i+1}] = c_i$
II: $A[t_{i+1}] \neq c_i$ and $t_{i+1} = 2r_i + 1$
III: $A[t_{i+1}] \neq c_i$ and $t_{i+1} \neq 2r_i + 1$

Proof of part (a) of the loop invariant:
case I $\vee$ II:
$r_{i+1} = r_i + 1$ (++r for I,II)
$t_{i+1} = t_i + 1$ (++t, as noted above)
$\leq 2r_i + 1$ (ind. hyp.)
$= 2(r_{i+1} - 1) + 1$ (first line)
$= 2r_{i+1} - 1$
$\leq 2r_{i+1}$
case III:
$r_{i+1} = r_i$ (III, no ++r)
$t_{i+1} = t_i + 1$
$\leq 2r_i + 1$
$\leq 2r_i$ (III)
$= 2r_{i+1}$

Point out that since we used $t_{i+1} \leq 2r_i + 1$ in both cases, we should shorten the proof by putting that before both of them.

Some preliminary gathering of results, perhaps by cases, might also save some repetition below (perhaps between (a), (b) and (c), perhaps between I, II III). You (TA) don't have to do that; instead point out to the students that once we've worked out a proof we should look it over.

Proof of part (b) of the loop invariant:
case I:
$c_{i+1} = c_i$
occur(c_{i+1}, $A[1..t_{i+1}]$)
= occur(c_i, $A[1..t_i + 1]$)
$\leq$ occur(c_i, $A[1..t_i]$) + 1 (actually = by I)
$\leq r_i + 1$ (ind. hyp.)
$= r_{i+1}$ (I)
case II:
$c_{i+1} = A[t_{i+1}] \neq c_i$
occur(c_{i+1}, $A[1..t_{i+1}]$)
$\leq$ occur(c_{i+1}, $A[1..t_i]$) + 1 (actually = by II)
$\leq t_i - r_i + 1$ (ind. hyp.(c))
$= (t_{i+1} - 1) - r_i + 1$
$= t_{i+1} - r_i$
$= 2r_i + 1 - r_i$ (II)
$= r_i + 1$
$= r_{i+1}$ (II)

case III:

$c_{i+1} = c_i$

occur(c_{i+1}, $A[1..t_{i+1}]$)

= occur(c_i, $A[1..t_i + 1]$)

= occur(c_i, $A[1..t_i]$) + occur(c_i, $A[t_{i+1},t_{i+1}]$)

= occur(c_i, $A[1..t_i]$) (III)

$\le r_i$ (ind. hyp.)

$= r_{i+1}$ (III)

Proof of part (c) of the loop invariant:

I'll be a little less symbolic/detailed here.

case I:

$A[t_{i+1}] = c_i$, so no new occurrences of $v \ne c_{i+1} = c_i$,

so occur $\le t_i - r_i$

$= (t_{i+1} - 1) - (r_{i+1} - 1)$

$= t_{i+1} - r_{i+1}$.

case II:

$c_{i+1} = A[t_{i+1}] \ne c_i$

Suppose $v \ne c_{i+1}$.

No new occurrence.

Case $v = c_i$:

occur $\le r_i$ by ind. hyp.(b)

$t_{i+1} - r_{i+1} = 2r_i + 1 - (r_i + 1)$ (by II) $= r_i$

Case $v \ne c_i$:

occur $\le t_i - r_i = t_{i+1} - r_{i+1}$.

case III:

c doesn't change; at most one new occurrence of a $v \ne c_{i+1}$; so

occur $\le (t_{i+1} - 1) - r_{i+1} + 1 = t_{i+1} - r_{i+1}$.