

Data Manipulation

Much of this course and lots of the lecture notes were inspired by or derived from Brown University's CS931. Our thanks go to Prof Shriram Krishnamurthi and Hammurabi Mendes for their kind permission in allowing us to use their materials.

From last time

- We compared all legislators to Starry Lee
- Use a score to compute their level of similarity to Starry.
- Ranked them by similarity to Starry.

Potential Problem

- Is Starry the best person to compare against?
- For example, let's consider Christopher Chung (鍾樹根)
 - Usually one of the most conservative
 - But — he will be ranked at some distance away from Starry Lee, and therefore will appear more pan-democratic!

An Analogous Example

- Order the stations along the East Rail line from Hunghom to Lowu
- But I'm not going to give you the names of the stations
- I'm only going to give you the distances (travel time) between the stations, and an example station
- Can you do it?



Distance (travel time) Table

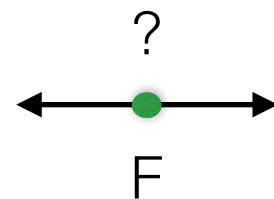
	Kowloon Tong	A	B	C	D	E	F	G	H	I	J	K
Kowloon Tong	0	26	9	4	12	6	16	8	45	24	31	34
A	26	0	17	30	14	20	10	34	19	2	5	8
B	9	17	0	13	3	3	7	17	36	15	22	25
C	4	30	13	0	16	10	20	4	49	28	35	38
D	12	14	3	16	0	6	4	20	33	12	19	22
E	6	20	3	10	6	0	10	14	39	18	25	28
F	16	10	7	20	4	10	0	24	29	8	15	18
G	8	34	17	4	20	14	24	0	53	32	39	42
H	45	19	36	49	33	39	29	53	0	21	14	11
I	24	2	15	28	12	18	8	32	21	0	7	10
J	31	5	22	35	19	25	15	39	14	7	0	3
K	34	8	25	38	22	28	18	42	11	10	3	0

Why is this not going to work?

Our Problem

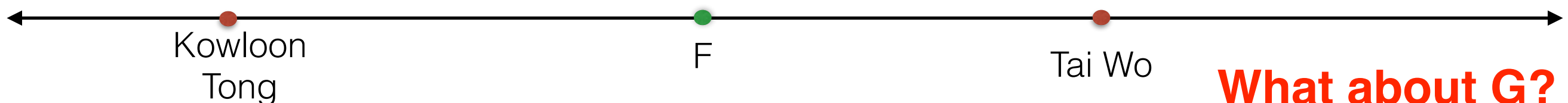
- Our example station, Kowloon Tong, is not at the end of the East Rail line.
 - Hunghom and Mongkok East are at some distance from Kowloon Tong
 - So are Tai Wai and Shatin.
 - But in a different direction
- Solution:
 - Use two example stations at different ends of the line (e.g. Tai Wo)
 - Hunghom will be at some distance from Kowloon Tong, and it will be at a larger distance from Tai Po.

	Kowloon Tong	Tai Wo	B	C	D	E	F	G	H	I	J	K
Kowloon Tong	0	26	9	4	12	6	16	8	45	24	31	34
Tai Wo	26	0	17	30	14	20	10	34	19	2	5	8
B	9	17	0	13	3	3	7	17	36	15	22	25
C	4	30	13	0	16	10	20	4	49	28	35	38
D	12	14	3	16	0	6	4	20	33	12	19	22
E	6	20	3	10	6	0	10	14	39	18	25	28
F	16	10	7	20	4	10	0	24	29	8	15	18
G	8	34	17	4	20	14	24	0	53	32	39	42
H	45	19	36	49	33	39	29	53	0	21	14	11
I	24	2	15	28	12	18	8	32	21	0	7	10
J	31	5	22	35	19	25	15	39	14	7	0	3
K	34	8	25	38	22	28	18	42	11	10	3	0



	Kowloon Tong	Tai Wo	B	C	D	E	F	G	H	I	J	K
Kowloon Tong	0	26	9	4	12	6	16	8	45	24	31	34
Tai Wo	26	0	17	30	14	20	10	34	19	2	5	8
B	9	17	0	13	3	3	7	17	36	15	22	25
C	4	30	13	0	16	10	20	4	49	28	35	38
D	12	14	3	16	0	6	4	20	33	12	19	22
E	6	20	3	10	6	0	10	14	39	18	25	28
F	16	10	7	20	4	10	0	24	29	8	15	18
G	8	34	17	4	20	14	24	0	53	32	39	42
H	45	19	36	49	33	39	29	53	0	21	14	11
I	24	2	15	28	12	18	8	32	21	0	7	10
J	31	5	22	35	19	25	15	39	14	7	0	3
K	34	8	25	38	22	28	18	42	11	10	3	0

Since Kowloon Tong - Tai Wo takes 26 minutes, Kowloon Tong to F takes 16 minutes, and Tai Wo to A takes 10 minutes, therefore, F must be between Kowloon Tong and Tai Wo



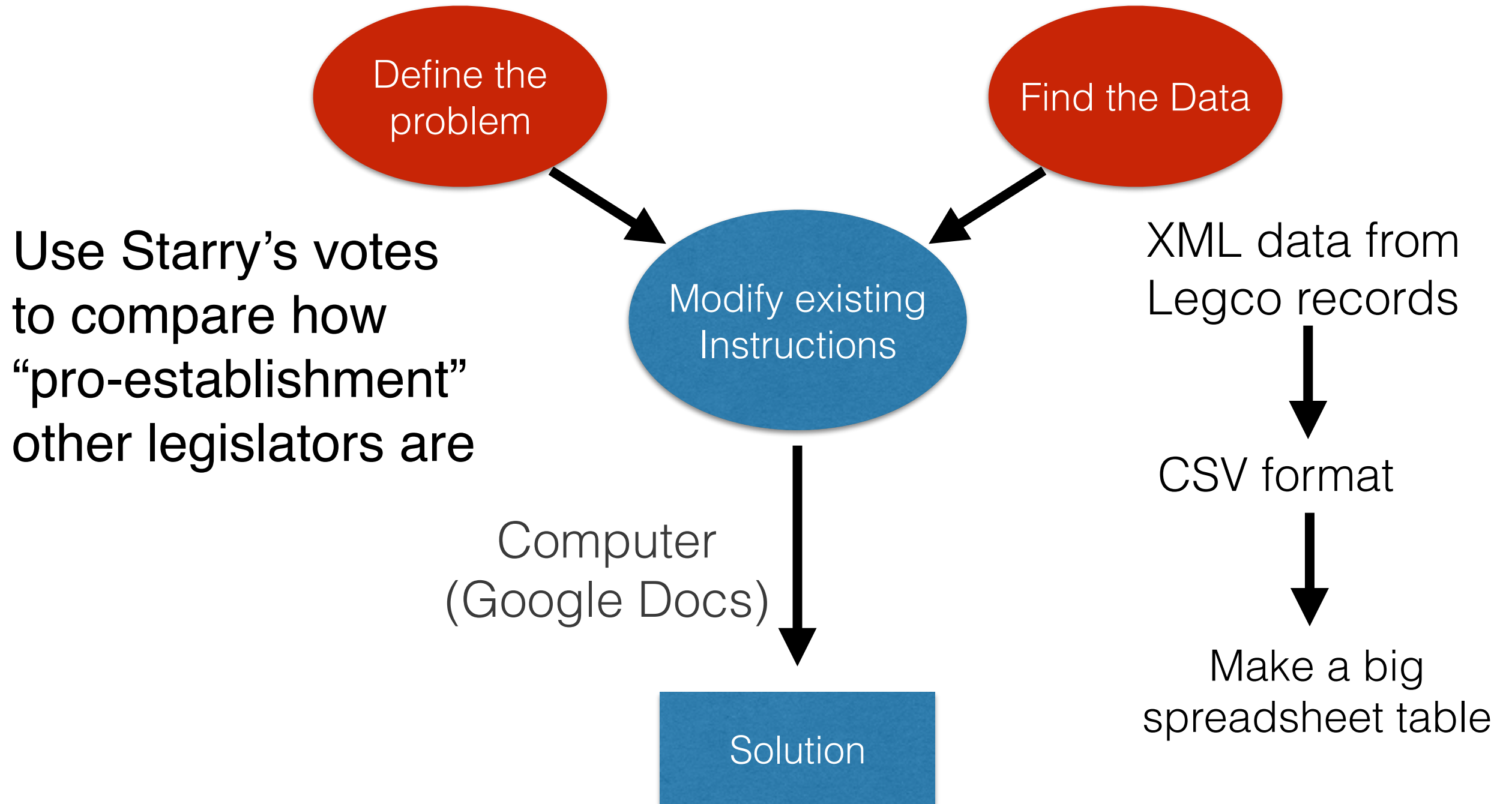
	Kowloon Tong	Tai Wo	B	C	D	E	F	G	H	I	J	K
Kowloon Tong	0	26	9	4	12	6	16	8	45	24	31	34
Tai Wo	26	0	17	30	14	20	10	34	19	2	5	8
B (Shatin)	9	17	0	13	3	3	7	17	36	15	22	25
C (MK East)	4	30	13	0	16	10	20	4	49	28	35	38
D (Fo Tan)	12	14	3	16	0	6	4	20	33	12	19	22
E (Tai Wai)	6	20	3	10	6	0	10	14	39	18	25	28
F (Univ)	16	10	7	20	4	10	0	24	29	8	15	18
G (H.H.)	8	34	17	4	20	14	24	0	53	32	39	42
H (Lowu)	45	19	36	49	33	39	29	53	0	21	14	11
I (Tai Po)	24	2	15	28	12	18	8	32	21	0	7	10
J (Fanling)	31	5	22	35	19	25	15	39	14	7	0	3
K (S.S.)	34	8	25	38	22	28	18	42	11	10	3	0



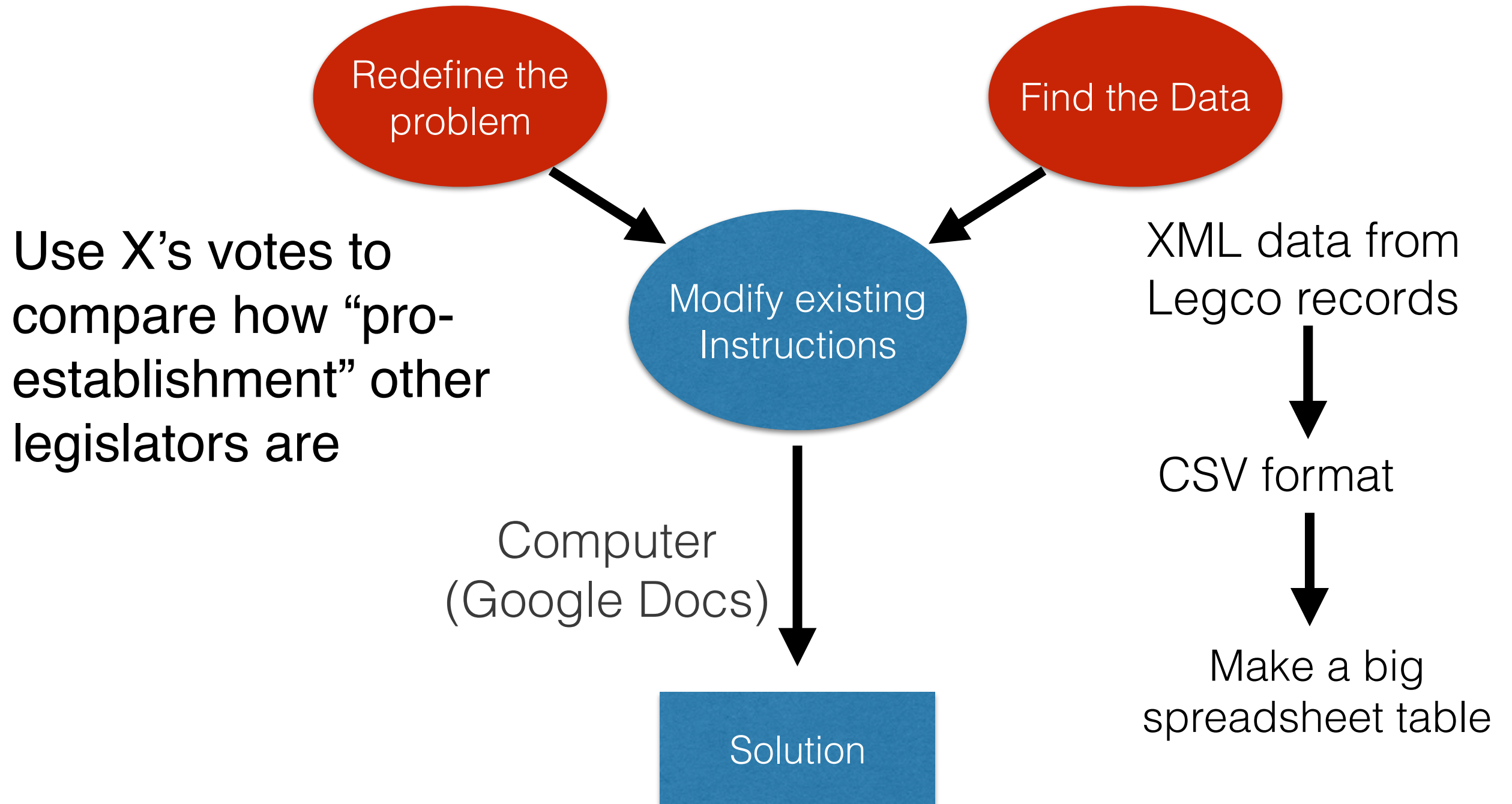
What does this mean for our problem?

- We do not need to pick the most “pro-establishment” or the most “pan-democratic” legislator in order for us to be able to do our ranking.
- BUT — we need two examples to compare against.
 - Either we modify our Starry table by hand to compare against another guy...
 - Or, we somehow *generalise* it so that we can compare against *any* other legislator.

What we had



New Problem

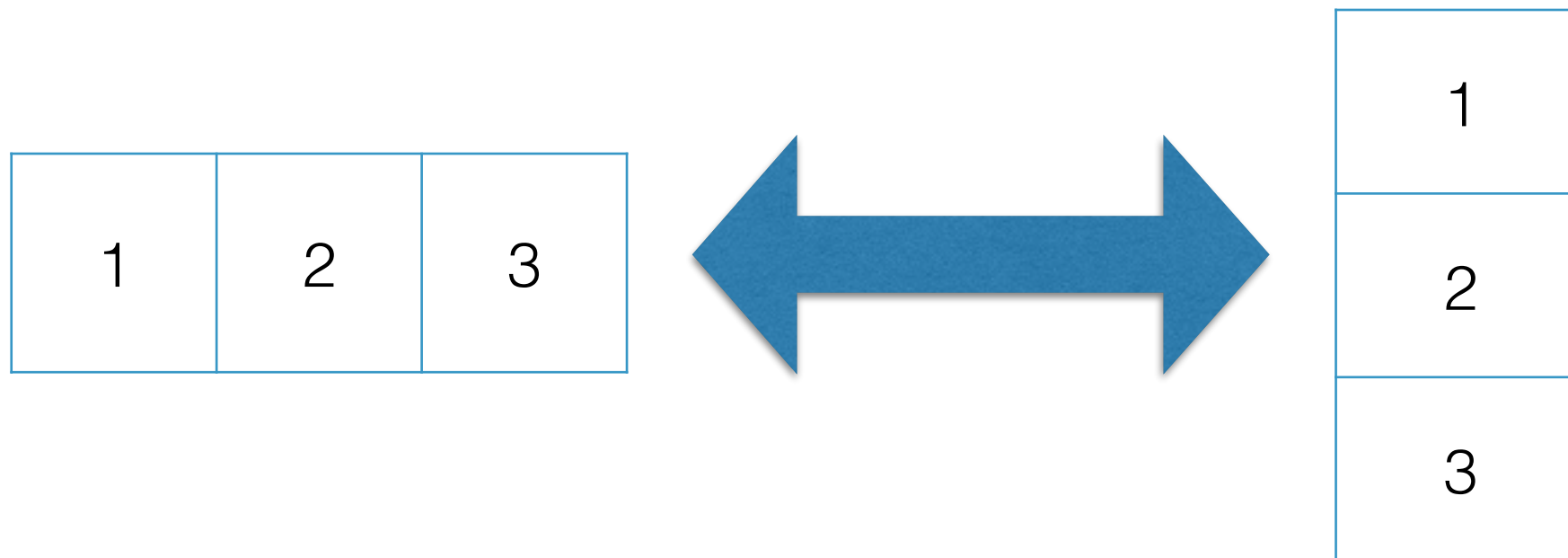


Activity 1-3

- Learn new spreadsheet techniques
- Rank relative to any legislator, not just Starry
- Task 2 will make a nice spreadsheet that can be used by others.

TRANSPOSE

- Makes a “row” into a “column” and vice versa
- Also works with ranges (2x3 becomes 3x2)



Reminder about **OFFSET**

- Offset: returns a reference that is a specified number of rows and columns away from a base *range* (or *cell*)
- `offset(reference, rows, cols)`

Activity 1-3

- Note how we broke this task down into little pieces
- Nice formatting (colors, etc) is useful.
- Validation is also useful.
- This spreadsheet could now conceivably be shared with other people as a useful tool.
- It's a *generalisation* of the old compare-with-Starry spreadsheet.
 - You can imagine how troublesome it would have been to change the old spreadsheet to compare with a new legislator!

Rankings

- Who is least like Starry?
- What are the rankings relative to that person?
- How do they compare to the Starry rankings?
- Rank Starry on your screen; get your partner to reverse-rank vs Starry's "opposite" and compare.

Next step: Compare
all orderings