

DATA TO GRAPHS AND BACK: SECONDARY TEACHERS' REASONING ABOUT THE AESTHETIC MAPPINGS THAT LINK DATA AND VISUALIZATIONS

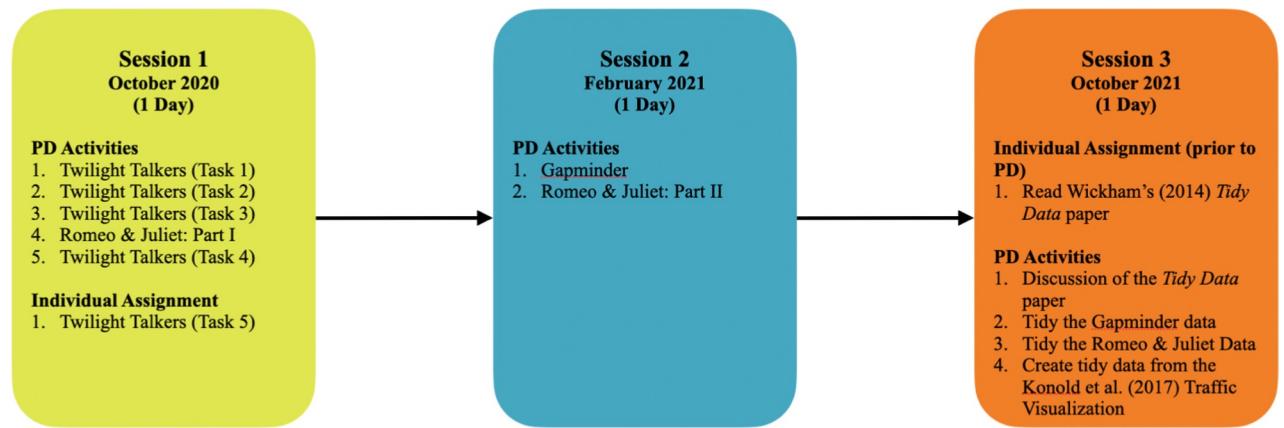
Chelsey Legacy, V.N.Vimal Rao, Andrew Zieffler, Robert delMas

Our Study



College in the Schools (CIS) Teachers ($n = 14$)

- Minimal coursework in statistics
- Bachelor's or Master's degree in Mathematics or Mathematics Education
- Some previously taught Advanced Placement (AP) Statistics
- Teaching CATALST curriculum for 1–8 years



Andrew Zieffler & Michael D. Huberty (2015) A Catalyst for Change in the High School Math Curriculum, CHANCE, 28:3, 44-49, DOI: 10.1080/09332480.2015.1099365

Research Questions

To what extent can secondary mathematics teachers (experienced in teaching statistics):

- use multivariate data to create a visualization that allows them to make sense of the potential multivariate relationships?
- reason from a data visualization depicting multivariate relationships to the raw data used to create the visualization?
- produce tidy data from a data visualization depicting multivariate relationships?

Research Question 1

To what extent can secondary mathematics teachers (experienced in teaching statistics) **use multivariate data to create a visualization that allows them to make sense of the potential multivariate relationships?**

Multivariate Data Table



Multivariate Visualization

Twilight Talkers Parts I, II, and III

Prompt: Work as a group to create a visualization that represents the interactions presented in the data table(s).

Character:	Interacts with:
Alice	Bella, Carlisle, Edward, Laurent
Bella	Alice, Carlisle, Charlie, Edward, Jacob, Jessica, Laurent, Mike, Sam, Tyler
Carlisle	Alice, Bella, Charlie, Edward, Laurent, Tyler
Charlie	Bella, Carlisle, Edward, Jacob, Mike, Tyler
Edward	Alice, Bella, Carlisle, Charlie, Jessica, Laurent, Mike
Jacob	Bella, Charlie, Jessica, Mike, Sam
Jessica	Bella, Edward, Jacob, Mike, Sam
Laurent	Alice, Bella, Carlisle, Edward
Mike	Bella, Charlie, Edward, Jacob, Jessica, Sam
Sam	Bella, Jacob, Jessica, Mike
Tyler	Bella, Carlisle, Charlie

Part I Table of Interactions

Sample of Part II Table of Characteristics

Character:	Frequency of interactions:		
	A little	Some	A lot
Alice	Carlisle, Laurent	Bella, Edward	
Bella	Jacob, Laurent, Sam, Tyler	Alice, Carlisle, Charlie, Jessica, Mike	Edward

Part III Table of Characteristics

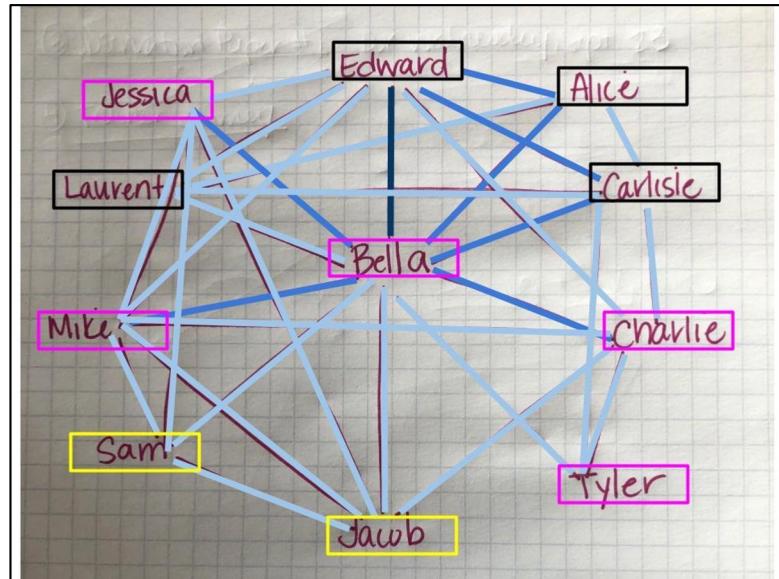
Species:	Character:
Human	Bella, Charlie, Jessica, Mike, Tyler
Vampire	Alice, Carlisle, Edward, Laurent
Werewolf	Jacob, Sam

Predictions

What do you think the teachers will come up with to display these interactions and character attributes?

Results: *Twilight Talkers Parts II & III*

Example 1

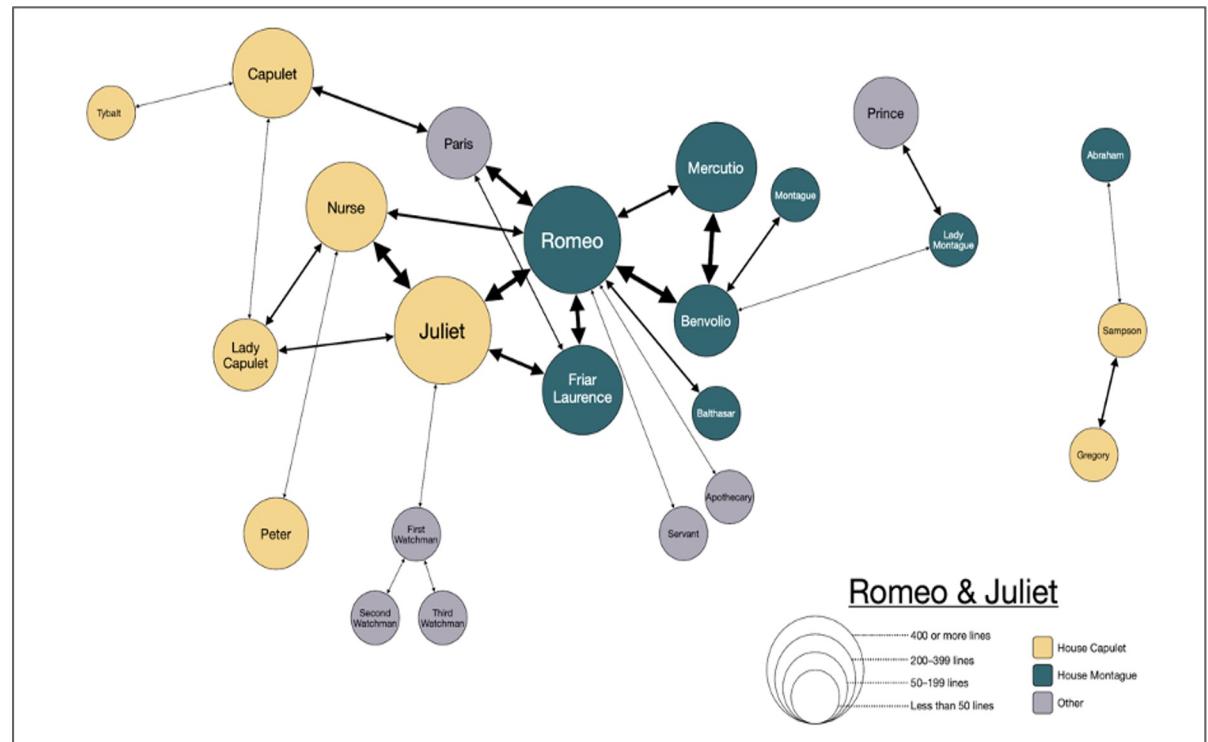


Example 2

Results: *Romeo & Juliet* Part I

- Formally introduced network graphs and aesthetic mappings
- Generally able to identify node and edge characteristics
- Lots of discussion about whether position, edge length, and font color were meaningful aesthetic mappings

"Romeo and Juliet are in the center for a reason, but I don't know if the other placements mean anything".



Research Question 2

To what extent can secondary mathematics teachers (experienced in teaching statistics) **reason from a data visualization depicting multivariate relationships to the raw data used to create the visualization?**

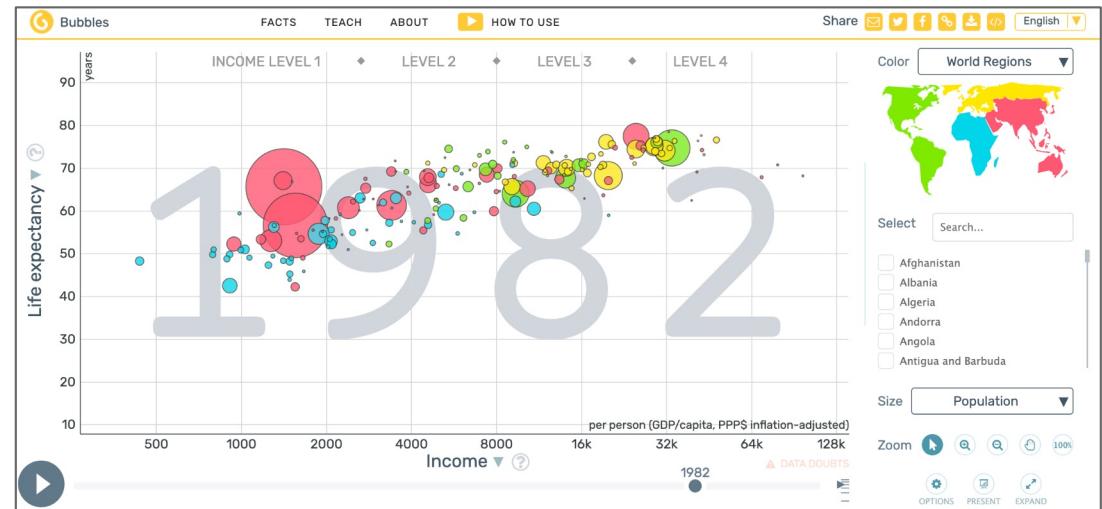
Multivariate Visualization



Multivariate Data Table

Gapminder Bubble Chart

Prompt: Provide the column names and the data—in a data table(s)—for the following countries from 1907, 1982, and 2019: Algeria, Andorra, Bolivia, Brazil, Georgia, Japan, Netherlands, Somalia, South Africa, and the United States



Predictions

What challenges do you think the teachers might face as they work to create a data table from this visualization?

Results: Gapminder Bubble Chart

Example 1

	A	B	C	D	E
1	Year	Location	Population (in mi	Life Expectancy	Income (Thousands \$)
2	1907	Africa	5.28M	29.5	1,910
3	1982	Africa	20.5M	65.1	11.4k
4	2019	Africa	43.1M	78.1	14k
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					

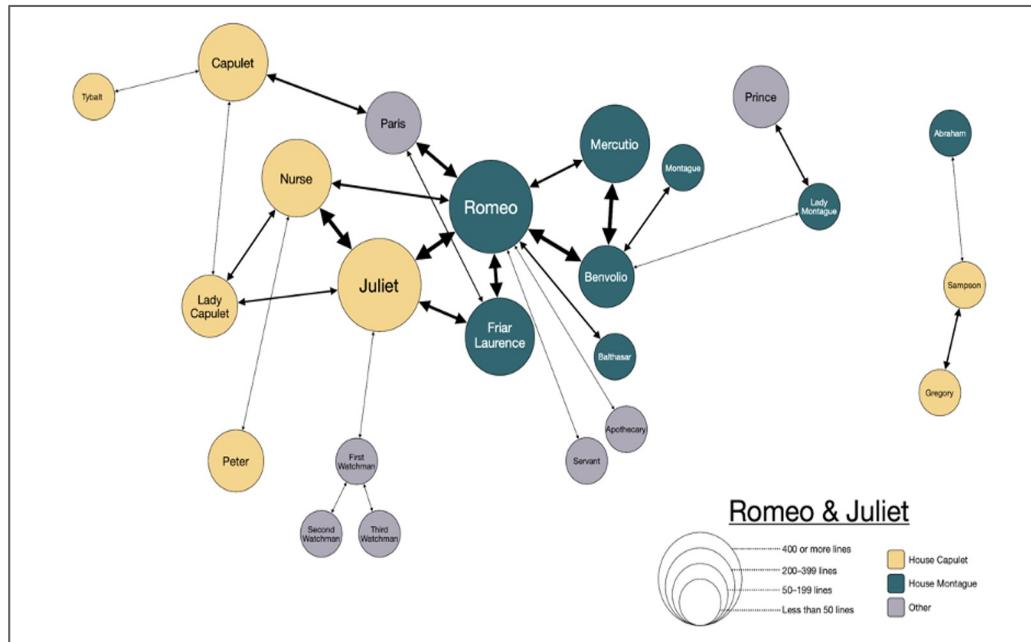
Example 2

Results: Gapminder Bubble Chart

When asked about creating multiple tables instead of one table, a participant replied:

“To me they are conceptually the same. We could have a data table for each country, year, we just happen to have it in one because we are so efficient. But I don’t think there is any difference. As far as having the data all in one place I don’t think there’s any advantage to having more than one [table].”

Romeo and Juliet: Part II



Prompt: Suppose you are asked to recreate this graph using a computer software program. The first thing you would need to do is enter the data into a data table (e.g., spreadsheet). Consider the data table(s) you would use to create this graph. Provide the column names and the data—in a data table(s)—for at least five cases in a data table.

Results: Romeo and Juliet Part II

Character	Number of Lines	House	Interaction
Romeo	400 +	Montague	Juliet - a lot
Romeo	400 +	Montague	Paris - a lot
Romeo	400 +	Montague	Servant - very little
Juliet	400 +	Capulet	Nurse - a lot
Juliet	400 +	Capulet	Lady Capulet - some

Example 1

	Fryer Laurence	Nurse	Juliet	Romeo	Paris			
Fryer Laurence	x	0	3	4	1			Interaction Key
Nurse	0	x	4	2	0		4	A lot of lines together
Juliet	3	4	x	4	0		3	Many lines together
Romeo	4	2	4	x	3		2	Some lines together
Paris	1	0	0	3	x		1	Very little lines together
							0	Less than 30 lines together
	House							
		Montegeau		Font Size 18	400+ Lines			
		Capulet		Font Size 14	200-399 Lines			
		Other		Font Size 10	50-199 Lines			
				Font 8	Less Than 50			

Example 2

Research Question 3

To what extent can secondary mathematics teachers (experienced in teaching statistics) **produce tidy data from a data visualization depicting multivariate relationships?**

Tidy Gapminder Data

Prompt: Is the data structure tidy?

Results: Tidy Gapminder Data

Example 1

Algeria	1907	Africa	Pop	Inc
Algeria	1982	Africa	Pop	Inc
Algeria	2019	Africa	Pop	Inc
Andorra	1907			
	1982			
	2019			
Bolivia	1907			
	1982			
	2019			

Example 2

Year	Country	World Region	Population	Income
1907	Algeria			
1982	Algeria			
2019	Algeria			
1907	Andorra			
1982	Andorra			

Romeo and Juliet Data

Prompt: Are the data tidy? If not, make them tidy.

	Fryer Laurence	Nurse	Juliet	Romeo	Paris				
Fryer Laurence	x	0	3	4	1				Interaction Key
Nurse	0	x	4	2	0		4		A lot of lines together
Juliet	3	4	x	4	0		3		Many lines together
Romeo	4	2	4	x	3		2		Some lines together
Paris	1	0	0	3	x		1		Very little lines together
	House								0
		Montegeau	Font Size 18		400+ Lines				
		Capulet	Font Size 14		200-399 Lines				
		Other	Font Size 10		50-199 Lines				
			Font 8		Less Than 50				

Results: Tidy Romeo and Juliet Data

House	Character	Interaction	Interaction #	Lines
Montague	F.L	NURSE	0	200-399
	F.L.	JULIET	3	
	F.L.	Romeo	4	
	F.L	Paris	1	
Capulet	Nurse	JULIET	4	50-199
		Romeo	2	
		Paris	0	
		F.L	0	

Results: Tidy Romeo and Juliet Data

House	Character	Interaction	Interaction #	LINES
Montague	F.L.	Nurse	0	200-399
	F.L.	Juliet	3	
	F.L.	Romeo	4	
	F.L.	Paris	1	
Capulet	Nurse	Juliet	4	50-199
		Romeo	2	
		Paris	0	
	F.L.		0	

Results: Tidy Romeo and Juliet Data

Character	Total # of Lines	Who Interacted	House of Character	# of Interact
Mit	400+	Nurse	Capulet	4
Juliet	400+	Fryer	Capulet	5
Fryer	200-399	Juliet	Montague	3
Juliet	400+	Juliet	Capulet	mult 0 15 null
Nurse	200-399	Fryer	Capulet	0 if * is a * intersection

If count = 4
then 10%

Results: Tidy Romeo and Juliet Data

Example 1

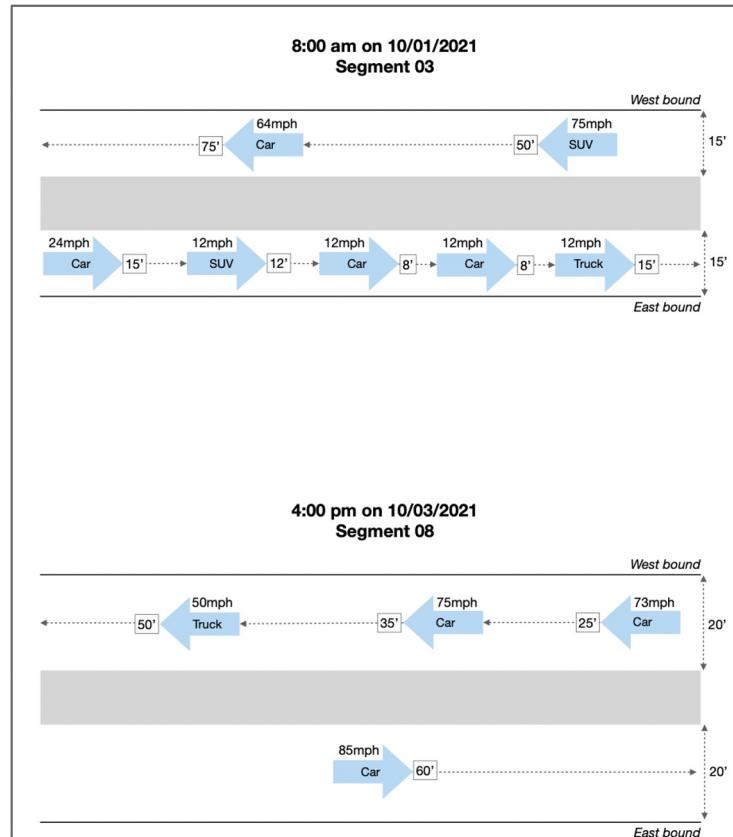
Character		
Character A	total #of lines	total #of character
Juliet	400+	Capulet
Romeo	100+	Montague
Interaction Table		
Character A		# of int.
Juliet	Romeo	4
Juliet	Nurse	4
Juliet		
Juliet		
Romeo	Nurse	
Character B		

Example 2

ID	House	Lines
1	Romeo	Montague 400+
2	Juliet	Capulet 200-300
3	:	50-100
4	:	<50
Interaction		
ID	Interaction with--	Amt
1	Juliet	A lot
1	Friar	MANY
1	:	Some
1	:	Few
2	:	None

Tidy Data from Traffic Visualization

Prompt: Your task is to create one or more organized (tidy) data sheets by recording the data from the snapshots. These data sheets should easily allow the city planners to add additional data to the sheets.



Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, 16(2), 191–212.

Results: Tidy Data from Traffic Visualization

- All groups realized they needed multiple tables
- Created separate tables in different ways
 - Most groups ($\frac{3}{4}$) split the information into two tables one about the car and its movements and the other about the segment/time
 - One group created 3 tables; with two containing information about the car and the third information about the segment/time
 - Different ways of creating IDs

Results: Tidy Data from Traffic Visualization

Car Data

Your task is to create one or more organized (tidy) data sheets by recording the data from the snapshots. These data sheets should easily allow the city planners to add additional data to the sheets.					
Car Id	Type	Speed	Direction	Distance in Front	Rd Id
1	Car	64	West	75	1
2	SUV	75	West	50	1
3	Car	24	East	15	1
4	SUV	12	E	12	1
5	Car	12	E	8	1
6	Car	12	"	8	1
7	Truck	12	E	15	1
8	Truck	50	W	50	2
9	Car	75	W	35	2
10	Car	73	W	25	2
11	Car	85	E	60	2

Rd Id	Segment	Date	Time	width	Vehicles
1	3	10/1/21	8:00am	15'	1-7
2	8	10/3/21	4:00pm	20'	8-11
3					

Segment Data

Results: Tidy Data from Traffic Visualization

Your task is to create one or more organized (tidy) data sheets by recording the data from the snapshots. These data sheets should easily allow the city planners to add additional data to the sheets.

Car Id	Type	Speed	Direction	Distance in Front	Rd Seg	Rd Id
1	Car	64	West	75	8:00	1
2	SUV	75	West	50	8:00	1
3	Car	24	East	15	8:00	1
4	SUV	12	E	12	8:00	1
5	Car	12	E	8	8:00	1
6	Car	12	"	8	8:00	1
7	Truck	12	E	15	8:00	1
8	Truck	50	W	50	8:00	2
9	Car	75	W	35	8:00	2
10	Car	73	W	25	8:00	2
11	Car	85	E	60	8:00	2

Rd Id	Segment	Date	Time	width	Vehicles
1	3	10/1/21	8:00am	15'	1-7
2	8	10/3/21	4:00pm	20'	8-11
3					

Results: Tidy Data from Traffic Visualization

vehicle ID	type	direction	vehicle ID	speed	segment ID	distance ahead
1	car	w	1	24	08 A	15
2	car	e	2	12	03 A	9
3	truck	w	2	12	03 A	9
4	truck	e	1	64	08 A	75
5	SUV	w	1	75	04 B	35
6	SUV	e	1	73	08 B	25

segment ID	segment	date	time	lanewidth	5	75	03 A	50
A	08	10/1	8a	15	6	12	03 A	12
B	09	10/3	4p	20	4	12	03 A	15
					3	50	04 B	50

Results: Tidy Data from Traffic Visualization

Car Data

vehicle ID	type	direction
1	car	w
2	car	e
3	truck	w
4	truck	e
5	SUV	w
6	SUV	e

Segment Data

segment ID	segment	date	time	lanewidth
A	03	10/1	8a	15
B	09	10/3	4p	20

Vehicle ID	Speed	segment ID	distance ahead
1	24	03 A	15
2	12	03 A	9
2	12	03 A	9
1	64	03 A	75
1	75	04 B	35
1	73	04 B	25
5	75	03 A	50
6	12	03 A	12
4	12	03 A	15
3	50	04 B	50

Car Data

Discussion of Results

- Intuitively able to go from data tables to visualization and identify visualization aesthetics
- Not as easily or consistently able to go from plot to data table
- Able to enact the first two tidy data principles with both the GapMinder and Konold et al. traffic snapshot data
 - **Each variable forms a column**
 - **Each observation forms a row**
- But showed progress with the third (**Each type of observational unit forms a table**) working on the Konold et al. traffic snapshot data

Questions/Comments