Research on Student Understanding of Data Organization

Datagames

Overview

What is a Data Game?

Students generate rich and interesting data when they play computer games — data that usually evaporate when the game ends.

In this project, we have created an environment in which data from computer games are available to student players for analysis. The games are designed such that successful performance depends on students being able to model data from the game.

Data From a Game

In a game called Chainsaw, the goal is to cut logs into pieces of a target length and to use as little fuel (time) as possible.

All of our games, including Chainsaw, generate data at two or more levels. Here, each cut piece has attributes such as length and accepted/rejected status. But there is also important data at the game level: game number, number of accepted pieces, fuel remaining.

A question faced early in the project was how to present and organize such hierarchical data for our students. Our prior observations were that students (and teachers) had considerable difficulty with much simpler data.

Research Questions and Participants

We designed the Traffic Problem to study how students at various ages organized multivariate, hierarchical data of the type that our games generate.

1. Could students come up with a method of recording complex data that maintained the critical information?
2. What kind of structures would they use to organize the data?

We have administered the Traffic Problem to participants ranging from middle school to university, some of these in an interview format using a think-aloud protocol.

Research Team

UMass Amherst

Cliff Konold, co-PI
konold@srri.umass.edu
Kosoom Kreetong, graduate research assistant

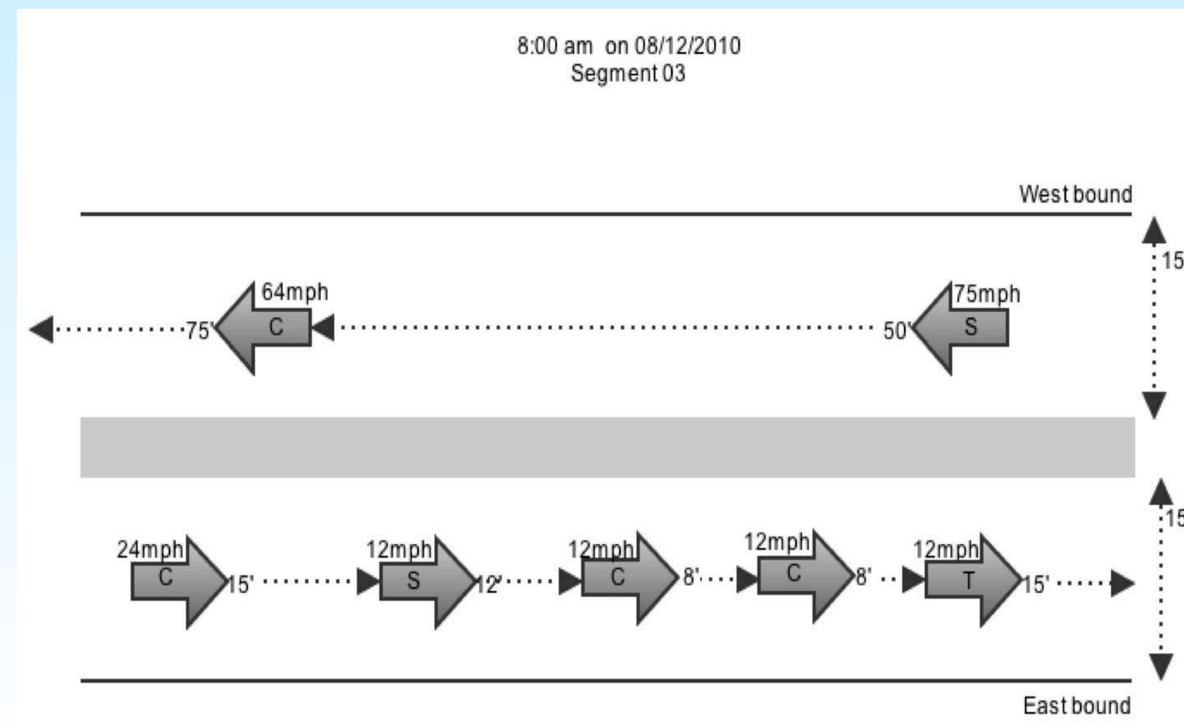
KCP Technologies

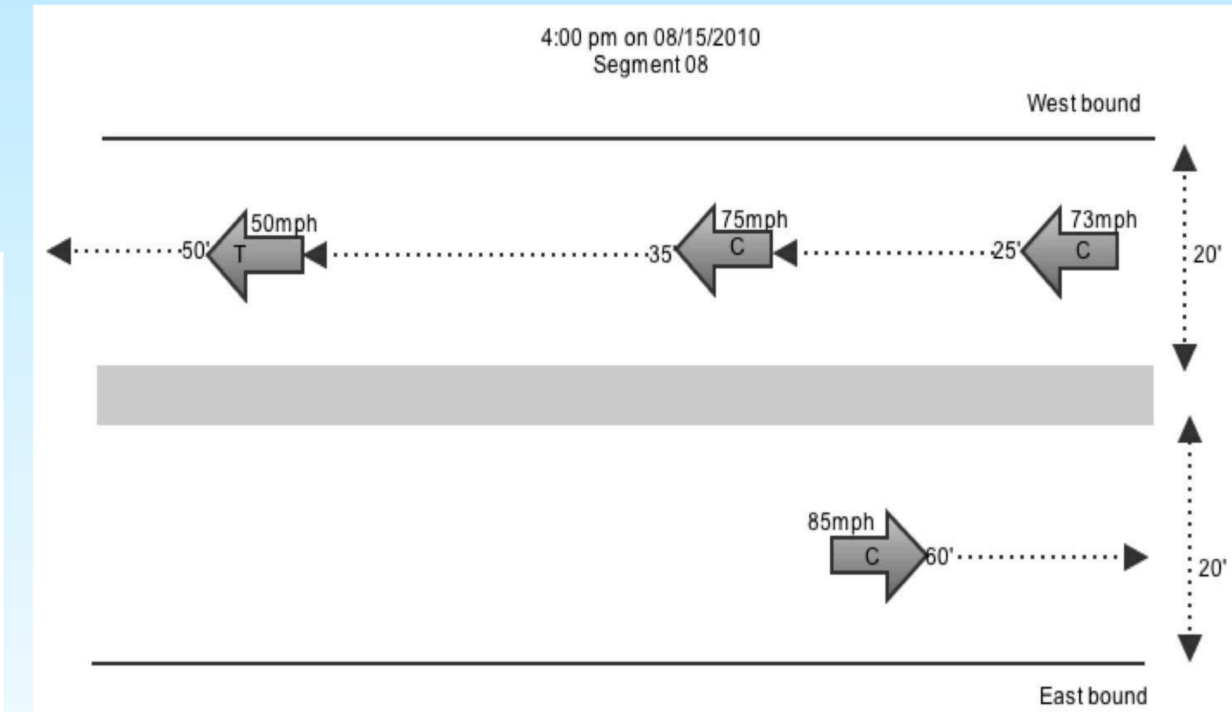
William Finzer, co-PI
bfinzer@kcptech.com
Rick Gaston, research and project manager
Vishakha Parvate, research and project manager
Jaimie Stevenson, research and project assistant

The Traffic Problem

These are snapshots that show the traffic along two road segments.

Suppose city planners are using snapshots like these to study local traffic, and they hire you to design the data form that will be used to record the information from these snapshots...



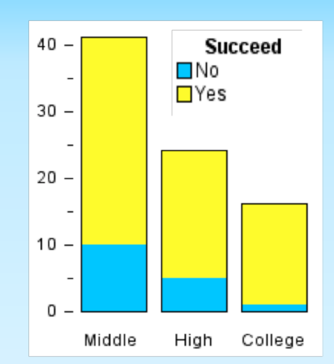


On a blank sheet of paper, record the data from these two snapshots. Your data sheet should not be a drawing of these snapshots. It should be an organized record of the data values of the snapshots....

Preliminary Findings

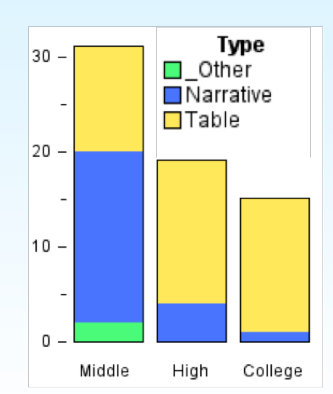
Success Rate

Surprising to us was the high success-rate even among middle school students. 76% of middle schoolers created a representation that could hold all the data while preserving the relations among them. Not surprisingly, the success rate trended upward over grades, reaching 94% at the college level (UC Berkeley undergrads taking a course on sampling).



Representations Used

Among those who successfully recorded the data, the predominant formats were tables and narratives (2 middle school students made graphs). Use of tables became more prevalent with higher grade levels.



Flat Tables

We categorized tables as either flat or hierarchical.

This is a flat table made by a 7th grader. Each row is a separate vehicle, with vehicle attributes listed along the top. Most statistical software programs, including *Fathom* and *TinkerPlots*, require that data be entered in this way.

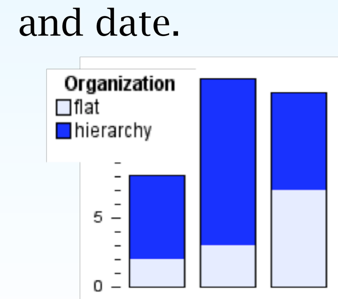
car Type	speed mph	Direction	Distence from car in frount	Time	Segment	DatE	road width
SUV	75mph	W	50ft	8:00am	3	8/12/2010	15'
car	64mph	W	75'	8:00am	3	8/12/2010	15'
car	24	E	15'	8:00am	3	8/12/2010	15'
SUV	12	E	12'	8:00am	3	8/12/2010	15'
car	12	E	8'	8:00am	3	8/12/2010	15'
car	12	E	8'	8:00am	3	8/12/2010	15'
truck	12	E	15'	8:00am	3	8/12/2010	15'
car	73	W	25'	4:00pm	8	8/15/2010	15'
car	75	W	35'	4:00pm	8	8/15/2010	15'
truck	50	W	50'	4:00pm	8	8/15/2010	15'
car	85	E	60'	4:00pm	8	8/15/2010	15'

Hierarchical Tables

This is a hierarchical table made by a college student.

Time & Date	Segment Number	Direction of traffic	Width of Lanes	Type of Vehicle	Speed (mph)	Distance to the vehicle in front (in ft.)
8:00am 8/12/2010	03	Westbound	15'	SUV	75	50'
				car	64	75'
8:00 am 8/12/2010	03	Eastbound	15'	car	24	15'
				SUV	12	12'
				car	12	8'
				car	12	8'
				truck	12	15'
4:00 pm 8/15/2010	08	Westbound	20'	car	73	25'
				car	75	35'
				truck	50	50'
4:00 pm 8/15/2010	08	Eastbound	20'	car	85	60'

Here, information about the vehicles, including type and speed, are again listed in a flat structure. But that flat structure is embedded in a larger structure that includes the information associated with the lanes, which is in turn imbedded in a larger structure that contains information about the snapshot's time and date.



Tables created by college students were evenly split between flat and hierarchical. For middle schoolers, however, the majority of tables were hierarchical.

Hierarchical Narratives

The majority of the narratives students made were structured as hierarchies. This is a hierarchical narrative made by a 7th grader.

8:00 AM on 8/12/2010
Segment 3
West bound:
Car 75 feet behind the next going 64mph
SUV 50ft behind the next going 75
East bound:
Car 15 ft behind going 24
SUV 12ft behind going 12
Car 8ft behind going 12
Car 8ft behind going 12
Truck 15ft behind going 12
4:00pm on 8/15/2010
Segment 08
West bound
Truck 50ft behind going 50
Car 35ft behind going 75
Car 25ft behind going 73
East bound
Car 60ft behind going 85

Summary

We had initially expected that flat structures would be used more frequently than hierarchical ones. But we've found just the opposite. We now believe that the flat structure for encoding the Traffic Problem requires a conceptual leap of sorts. Rather than thinking of lane width and direction as a feature of the road, the flat structure requires one to consider it as a feature of a vehicle.

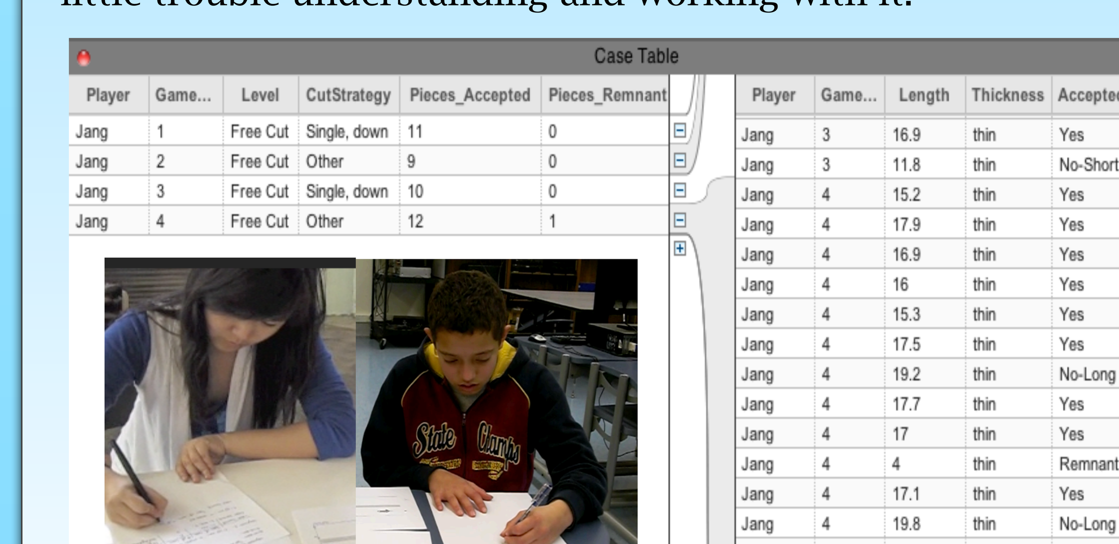
By contrast, the hierarchical structure recognizes different types of objects or events: the snapshot, taken on a particular time and data; lanes, with a width and direction; vehicles, with speed and type. The hierarchical narrative below highlights these different objects by enclosing each in their own box.

8:00 am on 08/12/2010
Segment 3
road
15 feet wide
east bound
1 car 24 mph 15 feet behind suv 1
1 suv 12 mph is 12 feet behind car 2
2 car 12 mph 8 feet behind car 3
3 car 12 mph 8 feet behind truck 1
1 truck 12 mph 15 feet behind something
west bound
1 suv 75 mph 50 feet behind car 1
1 car 64 mph 75 feet behind something

4:00 pm on 08/15/20
Segment 08
road
20 feet wide
east bound
1 car 85 mph 60 feet behind something
west bound
1 car 73 mph 25 feet behind car 2
2 car 75 mph 35 feet behind truck 1
1 truck 50 mph 50 feet behind something

Application to Software Design

Based in part on this research, we've designed a new hierarchical data table in DG. Our observations to date are that students have little trouble understanding and working with it.



We're continuing our analysis of the Traffic Problem including the interviews we conducted with some students where we can explore in more depth their constructive processes.

Acknowledgements

The Data Games project wishes to thank Galileo Academy of Science and Technology, Cupertino High School, KIPP Bridge, Bellarmine College Preparatory, Cornerstone Academy, Valley Christian High School, Amherst Regional High School, Amherst Regional Middle School, Four Rivers Charter School, UC Berkeley.

This material is based upon work supported by the National Science Foundation under:
KCP Technologies Award ID: 0918735
UMass Amherst Award ID: 0918653
Grant period: September 1, 2009 through August 31, 2012

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.