AN EARLY LOOK AT RICH LEARNING ANALYTICS: STATISTICS STUDENTS PLAYING "MARKOV"

Tim Erickson[1] and Mariel Triggs[2]
[1]Epistemological Engineering, Oakland, CA, USA
[2]Lick-Wilmerding High School, San Francisco, CA, USA
tim@eeps.com

In this paper we look at student work in a Data Game called "Markov," designed to give students experience with conditional probability. The online game records all student moves and strategy-making. We look at these "rich learning analytics" to see what patterns of student behavior we can see and what types of conclusions we might draw. We will see that these data are an intriguing window into student thinking and behavior, with implications for both instruction and research.

INTRODUCTION

We never know what really happens in the classroom. We see the results of student work but not the working. Even when students are doing an activity, and we circulate, eyes and ears wide open, we catch only snatches of student talk and glimpses of their tentative scribblings.

We may get video or audio recordings of individual, group, or class work—always instructive to us instructors. And if we're doing serious research, a painful period may follow, an eon of transcribing the media and developing coding systems so we can categorize the events, analyzing them for glimmers of learning and clues to student understanding. This kind of work is essential for research, but utterly impractical for day-to-day classroom decision-making.

This paper describes a middle ground available when students are doing an on-line activity: using "rich learning analytics." The basic idea is not new. Even Khan Academy has a "teacher dashboard" that tells instructors which exercises students have completed. But it does not tell them *how* the students did their work.

This paper describes the detailed data available for the game *Markov*, from Data Games (Finzer 2012). We will explore data from a class session to learn how students played the game and try to infer what they understand about conditional probability.

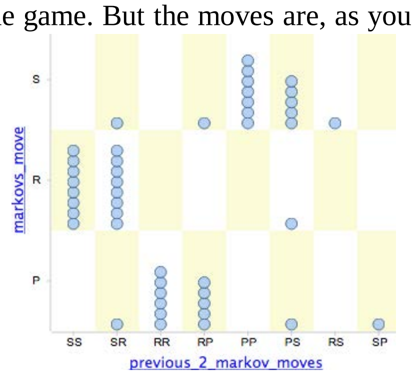
MARKOV, THE GAME

To understand the data, you need to know about the game. The very best way to learn it is to play it. Go to http://goo.gl/K7joio to play the game (press "login as guest" for access). You can also watch short videos about it at http://goo.gl/orW0wO (student introduction); http://goo.gl/IU9kIU (student advanced); and http://goo.gl/scuUgt (for the teacher).

In the game, you play a sequence of rounds of "rock-paper-scissors" (R P S) against an evil genius, Dr Markov. A dog, Madeline, sits on an elevator. Whenever Markov wins a round of the game, the elevator descends towards his basement laboratory. When *you* win a round, the elevator rises towards the surface—and freedom.

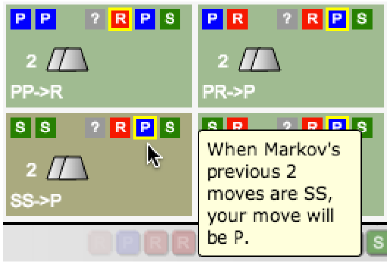
And the game is rigged: Markov wins all ties. So if you play randomly, you will lose two-thirds of the time. Madeline's only hope is for you to predict Markov's moves.

If his past dozen moves have been "R R P P S S R R P P S S" you might think that his next move will be R (rock)—and that therefore your next move should be P (paper). If Markov's moves were simply a repeating pattern, this would be enough to win the game. But the moves are, as you might expect, governed by a Markov process—so patterns do not repeat exactly.



Markov's sequence of moves appears on the screen and is recorded in a table. You can infer a pattern from that sequence, but the process is arduous. To make the decisions easier, you can make a graph that plots Markov's moves against his two previous moves. That graph shows the (empirical) conditional probabilities in the Markov process for each of the nine possible "antecedents" (SS, SR, RR, RP, etc.).

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org



The illustration shows such a graph. The first column, for example, shows that in this game, every time Markov has played two scissors in a row (SS) he has always followed it by playing paper (P). So you should play rock, as described above. It's not all certain, though; after PS, Markov usually plays scissors, but not always.

You can also make a *strategy* for the game. By this we mean that you can specify, for each antecedent, what your move would be. The illustration shows a few of the nine "strategy tiles" including the one that will play "paper" in response to Markov having moved SS. Back in the game, the system will use the strategy to play automatically if you wish.

USING ANALYTICS TO STUDY STUDENT GAME PLAY

Students can learn to read the graph, make strategies, and save the dog. But what actually happens? Fortunately, Data Games records every student on-line action. We can mine these "learning analytics" to see what students did.

For this paper, we recruited a class of 20 US high-school statistics students, aged 16–18, early in the academic year. The class session was 75 minutes. The beginning was taken up with introductions and playing a different Data Game (*Cart Weight*) in order to familiarize students with the interface. There was minimal instruction in how to play *Markov* except for occasional interactions with individual groups and a class-wide suggestion that it might be useful to make a graph.

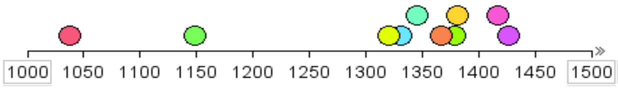


Figure 1. First-graph times (in seconds) for the ten groups. The activity started at 950.

This illustration is an example of the kind of information we can extract. It shows when the 10 groups first made graphs while they were playing *Markov*. We started playing *Markov* about 950 seconds into the data collection. Two groups made graphs spontaneously, while the other eight did so soon after the suggestion was made, about six minutes after they started playing (ca. 1310).

If we believe that looking at graphs is a good habit for students of data (and we do; see Finzer 2013), we see that this class does not yet have that habit. Later in the year, we would hope that more groups would make graphs spontaneously, and that the delay in creating graphs would be shorter.

We can also use analytics to look at a single group's work. In this class, each group produced several hundred records. We can look at these as a list and be fairly confident in broad inferences about what students are thinking and doing. For example, data show sequences of game turns interspersed with other events such as setting strategy. After setting strategies, the groups generally do better: the elevator goes up. This suggests that the students have figured something out. And when we look at the specific strategies they set, we see that they are generally good ones.

Reading the list is not as compelling as a graph, however. Figure 2 shows the height of the "elevator" over time for "Group 3."

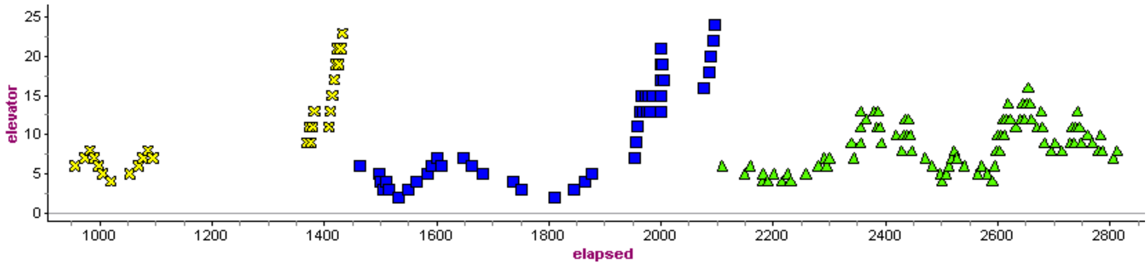


Figure 2. "Elevator" graph for one group (Group 3).

Each of the symbols represents one round; the symbols change when the group finishes a game and decides to change "level." (Each level has a different Markov pattern, so players need to change strategy.) This group played one game at each level. The group won the first and second games (cross and square)—the elevator reached the top at 25—but they never figured out the third game. In the first game, there is a gap of about 200 seconds; further investigation reveals that the group created a graph (aha!) and spent time setting a strategy. After the gap, the slope is steeper: they're letting the computer play automatically (and win) using the strategy they specified. In the second game, the "strategy gap," at about 1900, is shorter.

Different groups have very different patterns. "Group 8," in the next illustration, used autoplay much more extensively, completing a total of twelve games on four levels in the same time period. Their pattern for the first game was similar to Group 3's. Then, after quickly losing the first "square" game at about 1450, they paused to adjust their strategy and then won rather quickly.

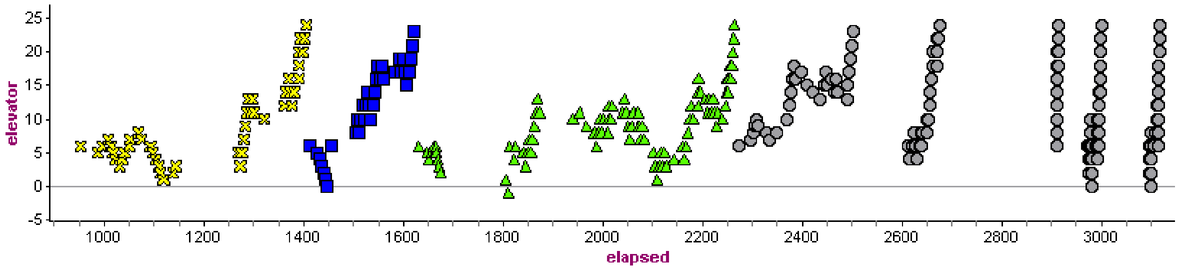


Figure 3. "Elevator" graph for Group 8.

Let's take a more detailed look at the second and third ("square") games. In Figure 4, symbols indicate the group's moves (left) and whether the moves were manual or automatic (right):

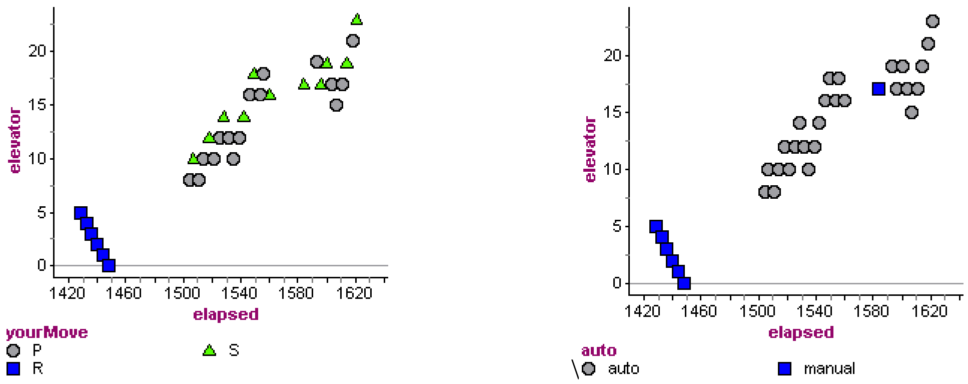
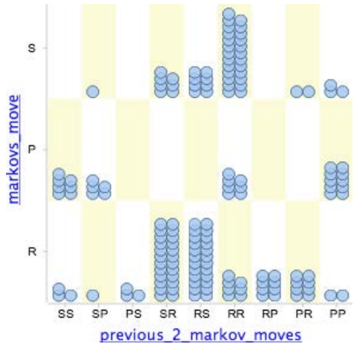


Figure 4. Second and third games from Group 8 in detail. At left, the group's moves; at right, whether they used "autoplay."

It looks as if Group 8 deliberately lost the first game, playing "Rock" quickly and repeatedly. A flurry of strategy-making (not shown on this graph) started at about 1470. It turns out this is a common pattern; some groups play quickly and lose, in order to amass enough data to make a better strategy. In this case, the group confirms this inference in their post-class writing: *When we started the game, we picked the same choice over and over so we could see what kind of patterns he was using.*

When we see that pattern in the data, can we infer that they're using this strategy? Not quite—but we can make such inferences more secure with additional research.

Now let's study the quality of moves. Looking back to Group 3's third game (*see* Figure 2)—the one they didn't complete—we can investigate what went wrong. Recall that when a group makes a move or defines a strategy, we know Markov's previous moves just as they did. So we can tell if their choice was a good one based on that historical data. Let's define **goodHistory** to be the number of rounds the players would have won if they had used that move every time that situation



had come up (i.e., whenever that set of Markov's previous two moves was the same). And we'll call the *total* number of such opportunities **totalHistory**.

Looking at the graph (right), for example, if Markov's previous two moves were "rock-rock," (RR), we should move "rock" (to crush the scissors we think Markov will choose). In that case, **goodHistory** will be 25 and **totalHistory** will be 39.

Basically, if **goodHistory** is more than half of **totalHistory**, Madeline will generally rise. The left-hand graph in Figure 5 shows the relationship between those two variables—and the "break-even" line—for Group 3's third game. (Notice that this graph is not a time series.) We've coded the points by Markov's two previous moves. You can see that the del-shaped points, indicating **prev2 = RR**, trail into the "bad" area below the line. The group had a bad strategy for that condition. Instead of playing "rock," they played "paper."

The right-hand graph *is* an elevator time series, with all the "RR" points highlighted (red). If the group had chosen a better strategy for RR, they would easily have won.

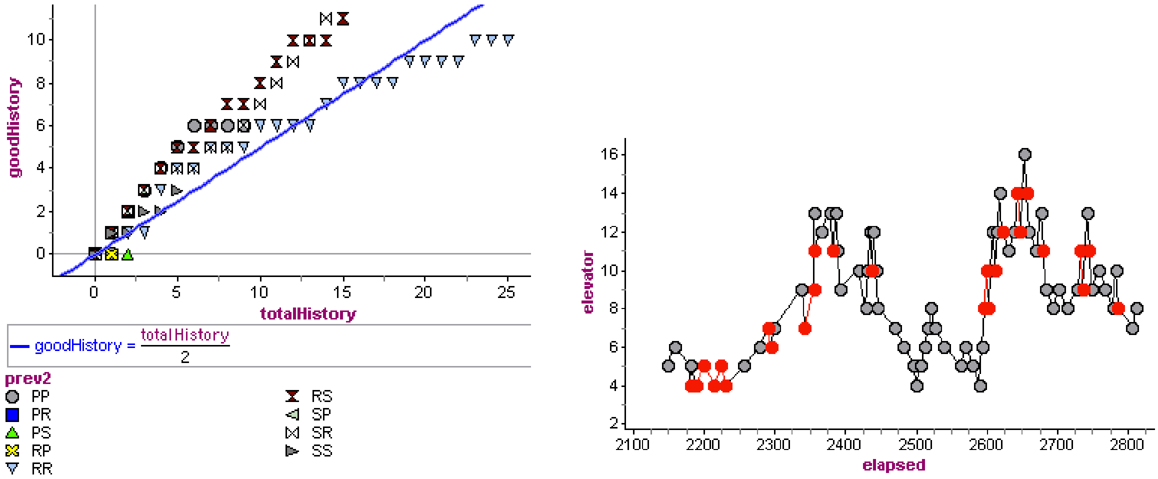


Figure 5. Different views of Group 3's unsuccessful third game.

How did Group 3 choose that bad strategy? Again, we can look back in the data—and discover that they had made a reasonable choice at the time. When they set that strategy tile to "paper," they had recently seen a long string of R's, and **goodHistory** : **totalHistory** was 5 : 8. *But they did not review their choice in light of new data.* Long strings of R's are rare and did not recur. By the time they gave up, that ratio was only 10 : 25. (On the other hand, data show that the plunge between 2450 and 2500 was primarily a string of bad luck.)

These have been only a few examples of what's possible. These learning analytics, especially when displayed in a graph, show us details about student work that are otherwise invisible. These rich data, of course, bring up new questions about what happened; yet we can often answer those questions with creative, nimble data analysis.

MARKOV AND CONDITIONAL PROBABILITY

Playing *Markov* does not teach conditional probability. It's background or reinforcement. An instructor can use it as a common referent when teaching or reviewing the more formal and opaque concepts associated with the topic. In the game itself, players need the underlying, conceptual idea of conditional probability to succeed. They must recognize the *condition*—Markov's previous two moves—and use the tools they have to determine, at least informally, the various conditional probabilities. In this game, that analysis takes place in a column in the graph.

Even inexperienced players understand this almost intuitively, and can come to understand the graph. Following Martignon & Krauss (2009), the graph shows individual cases rather than bars or pies. There are no hard formulas to apply. And the most useful and relevant graph is the default; the player doesn't need to configure it.

Even though the graph helps make it obvious, however, there are still challenges. For example, students have to change the way they look at Markov's pattern. It's easy to think of it as a sequence, e.g., "two rocks, then two papers, then two scissors, repeat." That is a compact, natural-language representation, and good for making predictions. But it breaks down when the pattern varies; then it's more useful and complete to think of it in terms of the nine conditions (defined by Markov's previous two moves) and the probabilities of each of three possible outcomes (Markov's next move). The graph in the game helps players make this shift from sequence to conditions-and-probabilities, and it's exactly the tool they need in order to specify a strategy. Sometimes, we can even see this transition in the analytics; the time gaps in Figures 2 and 3, when the groups made the graphs and then made strategies, are extremely suggestive.

Thus, analytics show that students are using data to make good decisions in a conditional-probability environment. Students win games, and their decisions (and speed in making them) improve with experience. But what do they really understand?

I am reluctant to draw firm conclusions. Does Group 8 understand the game better than Group 3? It's tempting to say so; they won more games, and actively generated the data they needed to strategize. But they also *lost* more games—and they were lucky. Furthermore, speed itself does not denote understanding. Some groups were, frankly, more devoted to saving Madeline than others. They were unwilling to "sacrifice" even a virtual dog in order to get data, so they chose their moves carefully—and more slowly. And it's an even bigger step to attribute success or failure at the game to understanding or confusion about conditional probability.

Nevertheless, these data can be one part of a comprehensive assessment strategy. They complement artifacts from student problem-solving and reflection. In addition, they help us detect student behaviors (e.g., not revisiting the strategy and paying the price) we might otherwise miss.

MORE POSSIBILITIES FROM ANALYTICS

What else should we do with such data?

These analytics may give us a window into hard-to-document student data practices. We've already mentioned the importance of making a graph early. In other settings, we would want to know what students made a graph *of*—and that information is available as well. In this game, when you change levels, there is a danger (despite on-screen warnings) of trying to analyze a new pattern using old data. We can see that mistake in the student records, and how the students recover from it. Perhaps we can see evidence of other important data-science habits-of-mind (Finzer 2013). We wonder how students clean data, how they organize it, what new variables they make, and so forth—and that is all accessible here.

Learning analytics can also help us be better teachers. We cannot be everywhere. If some students seem to be working productively, we might ignore them in favor of a group that is obviously having trouble. For example: in this class, although I was watching over students' shoulders, it escaped me that two of the ten groups hardly ever used the strategy feature or autoplay. For their work, I can analyze whether their moves were good (they were making good moves according to their graphs of Markov's historical data) but we have no evidence that they were confident enough to make their strategy explicit. Knowing this in retrospect, I can look harder at their work for evidence of understanding. And if I had known it in real time, I could have checked with them to see if they were even aware of the strategy and autoplay features.

HOW SHOULD WE ANALYZE LEARNING ANALYTICS?

Coarser-grained data such as that from online quizzes and skill practice tell us student scores and accomplishments. Elsewhere (Erickson 2013), I described how we designed Data Games leveling-up to ensure, for example, that anyone reaching the third level in the *Cart Weight* game has some understanding of linear functions. Here in *Markov*, anyone who completes two different levels with a preponderance of good moves probably knows how to interpret that graph—and that has some connection to understanding conditional probability. But the move-by-move microdata we're exploring here give us more detail—a more nuanced picture of a student's skills and understanding.

Which brings us to the problem of how we should use these data. First of all, taking *Markov* as an example, just looking at the "elevator" time series tells you a lot. You can see the different groups and how their profiles differ. It's a richer picture than a simple count of how many games a group won or lost.

We should immediately give teachers access to these data with useful representations at the level of the elevator graph. Such graphs help give a picture of what went on (or is going on right now) in class. We still might want to enhance that simple plot, for example, by showing when the group set strategy tiles; that would help explain gaps and help us infer understanding from subsequent speedy wins.

But it's hard to decide on the enhancements. As we said at the beginning, we're asking the data to stand in—if only partly—for our being present in the group. The elevator graph is clearly a step in that direction, but to get an even more detailed picture is harder. Could we automatically detect whether a group is sacrificing dogs for data? Can we flag a loss that occurred when a group failed to reset the data after changing levels? Can we easily tell whether a group simply got lucky with their strategy, and did not actually use their data well? All of these are possible, but difficult to get perfectly right. Moreover, we suspect that for every one of these interesting possibilities we want to study, there are two that we have not recognized. The data are rich and complex; it's still time for exploration.

So what's a teacher to do? At this point, I think the best approach is for teachers to analyze the data using the same tools the students use when they play the games. They will not have fancy, purpose-designed visualizations, but they can explore the data much as I have done in this paper, studying apparent anomalies and digging more deeply into what students may be thinking. The Data Games/CODAP system students use is now (January 2014) nearly capable of the analyses I have shown (I used *TinkerPlots* and *Fathom*); and teachers whose students use it will have an interest in being fluent in its use themselves. This possibility even allows the enticing metacognitive move of giving student records to students for analysis.

CONCLUSION: WHITHER RICH ANALYTICS

If a simple report of on-line achievements is "coarse" learning analytics, let's call this "rich" analytics. This report on the *Markov* game is an intriguing example. What's next?

As resources permit, I hope Data Games and its descendant, CODAP, make analytics like these more available. This will require thinking, testing, and tweaking of what data are recorded.

As researchers look for tools, I hope they incorporate rich learning analytics so we learn more about the validity of conclusions we draw from these data. Two paths immediately suggest themselves: more work connecting patterns in the data to our understanding of what students are thinking (as I did above with the student writing about sacrificing dogs for data); and the creation of other online environments whose analytic data can assess whether students learn less-proximal skills. That is, we know Markov students get good at *Markov*; can they transfer that to any other conditional-probability task?

Speaking as a teacher, I'm struck by the way these data let me "perch on the shoulders" of more students in my class; and even though they don't tell me definitively who understands what, rich analytics give me useful information, very quickly, that I never had before.

REFERENCES

Erickson, T. (2013). Designing Games for Understanding in a Data Analysis Environment. *Technology Innovations in Statistics Education*, *7*(2).

Finzer, W. (2012). *Fathom* [Computer Software]. Emeryville, CA: Key Curriculum.

Finzer, W. (2013). *Data Games* [Web Application Software]. http://ccssgames.com.

Finzer, W. (2013). The Data Science Education Dilemma. *Technology Innovations in Statistics Education*, *7*(2).

Konold, C., & Miller, C. (2012). *TinkerPlots* [Computer Software]. Emeryville, CA: Key Curriculum.

Martignon, L., & Krauss, S. (2009). Hands-on Activities for Fourth Graders: a Toolbox for Decision-Making and Reckoning with Risk. *International Electronic Journal of Mathematics Education*, *4*(3), 227–258.