



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Akshay Agrawal; Hemanth Kini	Project Number S1401
Project Title Plotting Muon Pathways as a Source of Randomness	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Truly random numbers are integral to various fields, ranging from cryptography to statistics. Many modern true random number generators (TRNGs), which extract random values from physical phenomena, are intrinsically biased. This project aims to create a novel TRNG by utilizing computer vision algorithms to analyze the position of muon pathways in a cloud chamber, since the emission of cosmic radiation is random.</p> <p>Methods/Materials A 99-percent isopropanol cloud chamber was constructed that allowed for the observance of muons. Muon streaks were photographed as RAW files. An OpenCV-based application was written to isolate the streaks by utilizing Gaussian blur and Canny edge detection algorithms. The same application was then used to extract the x and y coordinates of the midpoints of the streaks. The Cartesian coordinates were concatenated and used as input for ten tests from the NIST Statistical Test Suite.</p> <p>Results P-values greater than .01 were considered to imply randomness. Seven of the NIST tests indicated that 128-bit numbers generated from the x and y values were random, two indicated that 3000-bit numbers were random, while one indicated that 1000-bit numbers were random.</p> <p>Conclusions/Discussion The high success rate of the statistical tests strongly suggests that the data obtained is random. Since seven tests verified the randomness of 128-bit numbers, it can be concluded that the cloud chamber can be used to generate random 128-bit numbers. This project proves the viability of this cloud chamber as a novel, cost-effective TRNG while simultaneously providing insight into the random nature of cosmic radiation. Further research will focus on the automation of the entire process, including rendering the cloud-chamber self-sustainable and optimizing the computer vision algorithms for efficiency.</p>	
Summary Statement This project constructs a novel, reliable true random number generator through the utilization of muon pathways in a cloud chamber.	
Help Received Photography of the chamber was facilitated with the help of Gennadiy Magidin and Christophe Haubursin. The image processing procedure was advised by Dr. Ram Charan, Dr. Narendra Ahuja, and Ajay Agrawal. The procedure for the construction of the chamber was adapted from QuarkNet.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Macayla S. Ayers	Project Number S1402
Project Title Beginning Cartography	
Abstract Objectives/Goals The objective is to see if one could accurately measure the heights of various mountains using a transit and trigonometry. Methods/Materials Using a transit the angles of four different mountains were shot. Once the angles and distances (obtained from a GPS) were known, they were substituted into the formula $\text{tangent angle} = \frac{\text{the opposite leg}}{\text{the adjacent leg}}$ and the heights of the mountains were then calculated. After doing the math, the approximate heights found were then compared to the known altitudes determined by a GPS. Results The heights of all four mountains were calculated and only one mountain was accurate, having the same known and calculated heights, the other three mountains were off by a hundred to a few hundred feet. Conclusions/Discussion My conclusion is that with more accurate equipment one would be able to accurately measure the heights of mountains within a 95% accuracy.	
Summary Statement This project was conducted to discover if the height of a mountain could be accurately measured using a transit and trigonometry.	
Help Received My Dad taught me how to use the transit. My Mom helped me with the layout of my board. My uncle tutored me in trigonometry.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Ryan P. Batterman	Project Number S1403
Project Title Malware Identification by Instruction Level Code Analysis	
Abstract Objectives/Goals In this project, we created models of malware and goodware (normal software) using the composing assembly instructions of executables, and then we measured the effectiveness of these models at identifying malicious code. The objective is to determine the efficacy of this new method at distinguishing the two types of software. Methods/Materials 53514 malware programs were obtained from the Anubis research group, and 4115 goodware were collected from a Virtual Machine created for this project. We used the Naive Bayes, Kmeans, and Support Vector Machine algorithms to create models of malware and goodware, and then we determined the effectiveness of these classifiers at differentiating malicious code from normal code. Results The classifiers were effective at distinguishing malware from goodware. Conclusions/Discussion We successfully created models of malware and goodware that differentiate the two types of software using program assembly instructions. The results indicate that this method could likely be implemented in modern antivirus solutions.	
Summary Statement This project investigates the effectiveness of models that use program assembly instructions to differentiate malware from normal software.	
Help Received My teacher Dr. Durkee read my papers; My mentor Joshua Kroll taught me about machine learning and obtained the malware set.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Christine Chang	Project Number S1404
Project Title Beethoven Reborn: A Novel Approach Employing Markov Model to Synthesize Music in Great Composers' Styles	
Abstract Objectives/Goals The objective of this project is to use computers to learn music from great composers and generate new compositions that contain desirable characteristics as in the original music. Methods/Materials (I) Learning Phase: Implement Sequitur Algorithm to convert an original song from a sequence of notes to a sequence of motifs and learn the characteristics of the song, including motifs and motif's frequencies and the transition between consecutive motifs. (II) Composing Phase: Build a Markov table based on the information learned during the learning phase. Each (key, successor) pair maps to an entry in the Markov table. The computer program composes new music based on this Markov table. It randomly picks an initial key from the Markov table and randomly picks one of the successors, according to the learned probability distribution from the original song. The newly picked successor becomes the new key. This process continues until the number of motifs in the new song equals to the number of motifs in the original song. Composition techniques are applied to promote variations: with a tunable probability, the computer program elevates or lowers a motif by an octave, and it doubles or halves the duration of all notes in a motif. Results The combination of Sequitur Algorithm and Markov Model works effectively in learning original music and then in generating new music. Some of the computer synthesized songs sound beautiful, and most songs sound acceptable. Conclusions/Discussion Sequitur Algorithm is chosen to extract hierarchical information from songs. The hierarchical information is organized into a Markov table. New music is created based on the Markov table. Compositional techniques are applied to promote variations. A Java program is written to implement the algorithm. The program works efficiently and effectively generates beautiful new music with traces of the original music. Future research includes more intelligence in learning phase, for example, in the recognition and generation of transposition of any music interval, the processing of multi-notes and multi-parts, and the generation of more developed music structure. This approach of combining Sequitur Algorithm with Markov Model could be a powerful research method for applications such as text analysis and synthesis, genome scanning, or other applications that requires extracting and analyzing hierarchical information.	
Summary Statement This project identifies Sequitur Algorithm and Markov Model to be an excellent combination for computers to effectively learn the characteristics of input music and then to compose beautiful new music based on learned information.	
Help Received	



CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY

Name(s) Gwendolyn E. Chang	Project Number S1405
Project Title A New Method to Prove Euler's Equation by Using the Lagrange Mean Value Theorem	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Euler's Equation is beautiful formula combines five fundamental numbers in one and is generally considered to be the most elegant results in mathematics. Traditionally, we use four methods to prove it. We can use the Taylor series method which is a most popular way and can be found in our text books, the differential method, ordinary differential equations, or complex integrals. First, I enumerated the traditional methods. Then, I used the Lagrange Mean Value Theorem a new method to prove Euler's Equation. This method is clear and easy to understand. It is very helpful to understand both Euler's Equation and Lagrange mean value Theorem when learning calculus, especially for college and high school AP calculus courses.</p> <p>Methods/Materials My project was finding a new way to prove Euler's Equation by using the Lagrange Mean Value Theorem. First, I introduced and defined an auxiliary function $f(x) = \frac{e^{ix}}{\cos x + i \sin x}$, $I = (-\infty, \infty)$, and I used the flowing steps to prove the Euler's Equation</p> <ol style="list-style-type: none">1. Prove the Auxiliary Function $f(x) = \frac{e^{ix}}{\cos x + i \sin x}$ is defined before we use it by proving that the denominator $(\cos x + i \sin x)$ is not equal zero.2. Compare differentiability of real functions and complex functions and took the derivative of the function $f(x) = \frac{e^{ix}}{\cos x + i \sin x}$.3. Using the Lagrange Mean Value Theorem, #If $f'(x) = 0$ for all x in an interval (a, b), then $f(x)$ is a constant c on (a, b) and prove the constant $c=1$. <p>Results Since $f'(x) = 0$ and $f(x) = c$, when $x = 0$, $f(0) = 1$, therefore $f(x) = 1$ and $c = 1$. $f(x) = \frac{e^{ix}}{\cos x + i \sin x} = 1$, $[e^{ix}] = (\cos x + i \sin x)$ So when $x = \pi$, $\cos(\pi) = -1$ and $\sin(\pi) = 0$. We can conclude Euler's identity $e^{i\pi} + 1 = 0$ Euler's Equation has been proven.</p> <p>Conclusions/Discussion This project is pure and theoretical mathematics. It uses the Lagrange Mean Value Theorem to prove Euler's Equation because it is clear and easy to understand it. This method is very helpful to understanding both Euler's Equation and Lagrange mean value Theorem when learning calculus.</p>	
Summary Statement Using the Lagrange Mean Value Theorem to prove Euler's Equation	
Help Received None	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Jaston Epp; Samuel Taylor	Project Number S1406
Project Title Sorting Algorithms	
Abstract Objectives/Goals The objective of this experiment is to find which sorting algorithm is the fastest from five common algorithms. Our hypothesis was that mergesort would be the fastest as it takes advantage of it being easier to sort a short list than a long list and is able to break down the complexities of sorting thousands of numbers into many simple operations of comparing only two numbers. Methods/Materials In order to test our hypothesis, my partner and I created a program in Python that generates a list of 1,000 random integers in no particular order and then measures how fast each algorithm is able to sort that list from least to greatest. Materials include a computer running Windows and Python. Results In our 1,000 tests, the average sorting time for insertion sort was 0.000505785 seconds. The average sorting time for mergesort was 0.004430423 seconds. The average sorting time for heap sort was 0.005519048 seconds. The average sorting time for quicksort was 0.05777026 seconds. The average sorting time for Bubblesort was 0.106378149 seconds. Conclusions/Discussion Insertion sort turned out to be the fastest sorting algorithm. We believe this may be because of the extra time mergesort must have to merge lists.	
Summary Statement The project's purpose is to find the fastest sorting algorithm out of five common algorithms.	
Help Received CodeCodex was consulted to aid in writing the program. We learned Python from MIT OpenCourseWare.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Nikole Erickson; Alicia Springer	Project Number S1407
Project Title And the Winner Is...!	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals If a basketball or football team is winning at half time, then the majority of the teams will win the game overall.</p> <p>Methods/Materials Two thousand and fifty-six basketball and football game scores were collected, and the second and fourth quarter scores were examined. Eight hundred and seven of these scores were high school games (413 basketball, 394 football), 1,131 were college games (670 basketball, 461 football), and 118 were professional level games (58 basketball, 60 football). The scores were found in scorebooks provided by coaches, or were obtained from reliable online sources (espn.go.com and maxpreps.com). The 2nd and 4th quarter scores of each game were compared to determine if the teams winning at half time won the games overall. The scores were placed into either one of two categories; the 1st category consisted of games in which the team winning at half time won the overall game (win-win games), and the 2nd category consisted of games in which the team losing at half time won the overall game (lose-win games). Games in which the two teams were tied at half-time were excluded because of the nature of this project. Percentages of each of the categories were calculated, and graphs were made to illustrate the collected data.</p> <p>Results Out of the 2,056 basketball and football game scores collected, 1,734 games were categorized as win-win games, and 322 games were categorized as lose-win games. Eighty-four percent of the total number of games were win-win games and 16% were lose-win games. High school games accounted for 34% of the win-win games, college games accounted for 45%, and professional games accounted for 5%.</p> <p>Conclusions/Discussion Overall, the data did support the hypothesis. Eighty-four percent of the total numbers of games were games in which the teams winning at half time won the entire game. This 84% included 357 high school basketball games, 348 high school football games, 551 college basketball games, 384 college football games, 42 professional basketball games, and 52 professional football games. Sixteen percent of the total number of games did not fall under this category. This is important when considering whether basketball and football games need to have four quarters, or rather just two quarters. Because 16% is still a significant number, it is probable that basketball and football games will continue to have four quarters.</p>	
Summary Statement A comparison of 2nd and 4th quarter scores for high school, college, and professional level basketball and football games.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Abhijit S. Fnu	Project Number S1408
Project Title A Study and Analysis of Q-Learning Algorithm Parameters for Decision Making using a Developed Simulation Tool	
Abstract Objectives/Goals The objective of this research is to study Q-Learning Algorithm (QLA) and to develop simulation software in order to understand the optimal selection of parameters (learning rate and weight of future rewards) involved. Hypotheses: 1) If (a) both learning rate (alpha) and weight of future rewards (gamma) are defined to be 1; and/or (b) either alpha or gamma is set to 0, there is no learning involved. 2) To find the efficient path for reaching the goal, the optimal combination of alpha and gamma is 0.5 and 1, respectively. 3) If the sum of alpha and gamma equals 1, the average computation time (time to reach the goal) is constant, regardless of environment complexity. Methods/Materials QLA is a subset of reinforcement learning (RL) which falls under the Markov decision process (MDP). QLA pseudo-code forms the basis of this research. Other materials include a Windows-based laptop with 4GB RAM, C++ compiler, and an environment in which to test the learning agent. Results A virtual environment (i.e. simulation tool) was created from scratch with C++ compiler. An AI (Artificial Intelligence) agent was tested within the environment with discrete values of alpha and gamma. The computational time involved justifying the optimal path based on combined effect of defined values of alpha and gamma. 1) The first hypothesis was proven correct: (a) with alpha and gamma both set to 1, all states became goal states & (b) with either alpha or gamma set to 0, learning took infinite amount of time. 2) The second hypothesis was proven incorrect: the optimal combination of alpha and gamma was 0.9 and 1, respectively, as computation time was quickest with these values. 3) The third hypothesis is currently under study. Conclusions/Discussion MDP in AI domain is an unsupervised RL method involving mathematics and reasoning, computer algorithm, and software technology and emerging as an important area of interdisciplinary research as it has potential application in such areas as unmanned exploration, evolutionary research, and feature recognition. The research confirms that Q-Learning is a powerful technique that can be applied in the above areas.	
Summary Statement The Q-Learning Algorithm parameters, learning rate and weight of future rewards, were studied and analyzed in order to understand the effect of their optimal combined values by the use of a developed simulation software.	
Help Received My parents helped prepare the presentation board.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Joshua S. Goldwasser	Project Number S1409
Project Title An Exploration of Bases: Predicting Palindromic Patterns	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals My goal was to investigate the patterns of one's digits of perfect squares in different bases. Upon finding that the pattern was palindromic, I sought to explain the reason for that pattern and a means of calculating the particular pattern for any base.</p> <p>Results I proved that the patterns of one's digits of perfect squares are palindromes in any base, found a general formula for calculating the pattern in any base, and confirmed that the formula works in base 25. Along the way I also found that the patterns in bases that are divisible by four contain double palindromes.</p>	
Summary Statement I show why, in any base, the one's digits of perfect squares form a palindrome, and I find a general way of calculating its pattern.	
Help Received I designed, carried out, and wrote up this project myself. My advisor and my father made suggestions for editing and presenting it.	



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Sean S. Haas	Project Number S1410
Project Title Unbeatable PONG through Artificial Intelligence	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Since the dawn of the industrial revolution machines have worked side by side with humans, but they have not always been able to reason out and overcome human inconsistencies. The goal of this project was to make an artificially intelligent computer program to play the novel game, PONG. The computer program mimicked human thought processes, and its programmed speed and efficiency was tested against humans. My null hypothesis was that there would be no difference in the success of a computer controlled paddle and a human controlled paddle while playing the PONG game.</p> <p>Methods/Materials Three programs of PONG were written to determine the successes of both humans and a computer program while playing PONG. Human vs. wall and computer vs. wall games were used to gather baseline data, and a human vs. computer game was designed to compare relative successes in an actual competition. The computer used a modified slope-intercept equation to predict where the ball would intercept the wall, and moved its paddle to the point of intersection. The number of successful hits by each player was divided by the total number of possible hits; this rendered a percentage of successful hits and rates of success. I used the baseline tests as controls to compare to the results from the human versus computer game. These programs were written using Microsoft Visual C# Studio.</p> <p>Results All computer trials, both against a wall and against a human opponent, obtained a 100 percent success rate. Human success rates ranged from 62 to 85 percent in both the baseline tests and the actual game, with an overall average success rate of 73 percent.</p> <p>Conclusions/Discussion The null hypothesis that there would be no difference in the success rate of a computer programmed paddle versus a human controlled paddle while playing PONG was rejected ($P < .001$). The computer was significantly more successful than humans while playing PONG. My results demonstrated how a computer or machine can be programmed to carry out human tasks much more efficiently than a human. The results also demonstrate how computers can react logically and more efficiently than humans. The results also show that the use of computers to calculate predictions could easily be expanded to be used in assembly lines with on the fly adjustments and self troubleshooting capabilities to minimize the effects due to human error.</p>	
Summary Statement I wrote artificially intelligent computer programs to test against humans.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Sana Hadyeh	Project Number S1411
--------------------------------------	---------------------------------------

Project Title T.R.E. Method of Forestry
--

<p>Objectives/Goals</p> <p>My investigative question is which small sample surveying sampling method has the best representation of a larger whole population to manage the growth of trees; transect surveys, radial surveys, equilateral surveys, or all 3 combined? A forest is too large to count every tree. Sample surveys of smaller areas may need to accurately represent a much larger population area. My question examines what percentage of the larger population can be represented by smaller transect, radial, and equilateral survey samples. With previous knowledge about radials and transects, I decided to design a new method, called equilaterals. This method expands in the forest at a 90 degree angle.</p> <p>Abstract</p> <p>Methods/Materials</p> <p>I constructed 4 models of 24 scaled down #hectares#. Each hectare represents 2.4 acres or 10,000 sq. meters. I then labeled them. On model 1 I drew 2 transects, 2 radials, and 2 equilaterals. On model 2, I drew 6 transects. On model 3, I drew 6 radials. On model 4, I drew 6 equilaterals. Each representative sampling area was equal to 1 hectare in size. I randomly created a forest of 200 trees per model. I determined the mean average of each smaller sample, and compared them to means of the larger populations. I determined validity with a t-test in each case at a 95% confidence level.</p> <p>Results</p> <p>Percentage of small samples that met the T-test @ 95% confidence level.</p> <table><tr><td>Model 1: Combination of all three sampling methods</td><td>21.20%</td></tr><tr><td>Model 2: Transect sampling method</td><td>29.68%</td></tr><tr><td>Model 3: Radial sampling method</td><td>8.48%</td></tr><tr><td>Model 4: Equilateral sampling method</td><td>76.34%</td></tr></table> <p>Conclusions/Discussion</p> <p>Surprisingly, the equilateral methodology had by far, the best representation of the larger whole population at 76% accuracy. This is surprising as the surface area represented was equal in all sampling methods.</p>	Model 1: Combination of all three sampling methods	21.20%	Model 2: Transect sampling method	29.68%	Model 3: Radial sampling method	8.48%	Model 4: Equilateral sampling method	76.34%
Model 1: Combination of all three sampling methods	21.20%							
Model 2: Transect sampling method	29.68%							
Model 3: Radial sampling method	8.48%							
Model 4: Equilateral sampling method	76.34%							

Summary Statement This project mathematically examines the validity of three sampling methods used in field biology studies.
--

Help Received



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Johnny Ho	Project Number S1412
Project Title Post-Disaster Response Using a Novel Adaptive Object Recognition Algorithm on High Resolution Satellite Images	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of the project was to design a fast and accurate automated system for recognizing objects in satellite images. By locating tents, this automated system would then help NGOs locate displaced populations and allow them to use their resources more effectively in a post-disaster situation. This system is necessary because current manual methods are ineffective, and current algorithms do not work well on tents.</p> <p>Methods/Materials The experiment was carried out on a standard personal computer, and also utilized publicly available satellite imagery. The novel object recognition algorithm was divided into four components. First, images are compared by constructing feature histograms. Second, to improve accuracy, multiple filters (Edge detection, line detection, Gabor, Tamura, etc.) are combined when comparing images. Each of these filters contributes unique feature data to the algorithm. Third, a noise reduction algorithm is applied to the feature histograms, allowing for more accurate comparison of histograms. Fourth, simulated annealing is used as a learning algorithm to weight the strongest of the multiple filters. After running the system over the entirety of Haiti, the accuracy of the results was measured using an F score. The system was then compared with a publicly available set of manually-labeled tents, and also tested on various types of test data.</p> <p>Results The system was more effective than the manually-labelled data, as it was more accurate, efficient, and required less manual work. The final system produced an accuracy rate of 95.2%, which was 189% higher than that of the manually-labeled data. This multidimensional system was also determined to be 94% more accurate than any of the individual filters alone. In addition, the final system was similarly accurate when tested on various other types of data, such as trees or collapsed buildings.</p> <p>Conclusions/Discussion The automated system successfully recognized objects in satellite images, and was thoroughly tested in the process. The system was able to accurately and efficiently identify satellite images, surpassing a previous attempt at manually labeling tents. This shows that it is possible to more effectively locate displaced populations using automated systems and satellite images. Furthermore, this system can be extended to a large variety of other applications, such as deforestation or road detection.</p>	
Summary Statement To improve upon current manual methods, I constructed a novel adaptive object recognition algorithm for locating displaced populations using satellite images.	
Help Received Parents constantly supported me throughout the project; Brother (Tony Ho, Harvard class of 2014) guided me in the exhibition process.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Nadav Hollander	Project Number S1413
Project Title Developing a Computer Program That Effectively Mimics Human Creativity in Composing an Original Musical Melody	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The purpose of this research project was to develop a computer program that randomly generated musical melodies in a given key signature, and then to revise this program in a way that it would emulate the human creative process of music composition in its generation of musical melodies.</p> <p>Methods/Materials Two versions of the program were written, with the first version being a simple randomized melody generator, and the second version containing several algorithms I developed in order to mimic the human creative process of music composition. In version 1.0 of the Melody Generator, a "while" loop was set up to generate random fitting rhythm values in one function that would be transferred over to another function. This function would then match randomized note values within the given key to the rhythm values and output the results. In version 2.0 of the Melody Generator, four algorithms were developed to add musical coherency to the melodies produced: The Downbeat Algorithm, Tonic Algorithm, Leading Tone Algorithm, and Rhythm Repetition Algorithm. The resulting melodies of both versions were then tested and rated in a blind test on a 1-10 scale of coherency ten times for each given key.</p> <p>Results The melodies generated by version 2.0 were on average rated 27% higher than those generated by version 1.0. The mean coherency rating produced from version 1.0's melodies was a 3.4 out of 10, while the mean coherency rating produced from version 2.0's melodies was a 6.1 out of 10. Melodies produced by version 2.0 were far more coherent, logical, and musically pleasing than those produced by version 1.0.</p> <p>Conclusions/Discussion The data supported my hypothesis that, through the implementation of several AI algorithms that emulate the human musical composition process, the second, revised version of the music-generating-program would generate more coherent and pleasing melodies than those produced by the first version. Melodies produced by Version 2.0 of the program were rated an average 27% higher than melodies produced by Version 1.0, and, overall, sounded far more coherent, legible, and musically pleasing than the melodies produced by Version 1.0. It can be assumed that the development and implementation of the Tonic Algorithm, Downbeat Algorithm, Leading Tone Algorithm, and Rhythm Repetition Algorithm directly increased the ratings of Version 2.0 melodies.</p>	
Summary Statement This project centered around the development of a computer program that composes musical melodies in a way that emulates the human creative process of music composition.	
Help Received	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Paras J. Jain	Project Number S1414
Project Title SmartCheck: Innovating Credit Card Security through Smartphone Based Handshake Protocols, Fingerprinting, and Encryption	
Objectives/Goals I have a very strong motivation for my project. Five months ago, VistaPrint stole my parent's identity and made many fraudulent charges to the card weeks after the initial purchase. The problem with today's system is that information can be reused. Identity fraud is responsible for a \$221 Billion loss every year # my project saves that. My question was: How can credit card security be improved in online and in-store transactions without reducing consumer ease-of-use? My hypothesis for my question was: Credit card security can be improved without harming consumer ease-of-use by using smartphones to replace the physical card.	
Abstract Methods/Materials The first step was to determine the effectiveness of modern day encryption. I ran a test with the three most used hashing algorithms, MD5, SHA1 and CRC16/32 against my two algorithms, SCA and SCB. Five trials were conducted. My two algorithms also included information on the user making the request, making it mathematically at least 1,000x as difficult to break. I found that my SCB algorithm was practically unbreakable with modern hardware. I used a smartphone, a Secure Server Stack, a POS Thin Client, a fingerprint reader and a top-of-the-line test computer. I then had to create an implementation of my design. I made an iPhone application that would generate a verification code from purchase information and many pieces of user information. The hash is practically unbreakable so this information can't be accessed. The vendor sends this code to the server and if the code generated on the server matches, payment is authorized.	
Results The results of this were tremendously promising. It took 8.3hours and 12.7hours to crack CRC 16/32, respectively. It took 35.2hours to break MD5. It took 28.5hours to break SHA1 and 92.7 hours to break SCA. Over the period of 2weeks, SCB was unbroken and little progress had been completed. The tests were run in worst-case-scenario where the hypothetical hacker had access to the server, the code and the database.	
Conclusions/Discussion The practical applications of my research are huge. Primarily, the credit card industry could benefit from this. Also, business security and the encryption of military applications can be greatly secured as my SCB algorithm was practically unbreakable. I was able to break the SHA1 algorithm in 30hours # frightening as it is used heavily by the US military. This experiment was a fantastic success.	
Summary Statement This project aimed to find a way to improve the security of credit cards without reducing user ease-of-use; the results are promising as the encryption I created is unbreakable with current hardware and the implementation was secure.	
Help Received All of the project was done independently but I would like to thank: Family and Teacher helped motivate me through the project; Brother helped make board; Mother helped design of board; Sister helped proofread submission; Father inspired me to go into STEM	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Abraham P. Karplus	Project Number S1415
Project Title Set Phasors to Analyze Circuits	
Abstract Objectives/Goals My purpose with Set Phasors to Analyze Circuits was to create an electrical circuit analyzer, which is a computer program that can find the current and voltage of an electrical circuit (containing resistors, inductors, capacitors, voltage sources, and current sources). Note that no actual circuit is involved; rather, it is all modeled on the computer. Methods/Materials I wrote the program in the Python programming language with the help of the Numpy and PyGUI packages, for matrix math and a graphical user interface respectively. It uses the extended sparse tableau algorithm for finding voltage and current, and uses phasors, complex numbers representing a sinusoid, to allow it to handle alternating current with the same code as direct current. Results The program produces the correct results and can plot them on a graph or output as an RDB file for analysis or plotting with an external program. It has a graphical user interface, though still uses text entry for the circuit model input. It can "sweep" multiple AC frequencies to demonstrate the frequency response of the circuit. Conclusions/Discussion I successfully created a circuit analyzer, and learned a lot of programming skills and circuit theory. Note: My actual abstract in my report contains a lot more circuit theory, but the hypothesis-experiment-result formatting of this abstract submission form does not let me go into detail.	
Summary Statement I created an electrical circuit analyzer, a computer program that models a circuit to find its current and voltage.	
Help Received My dad, Kevin, helped me learn Python and circuit theory. My mom, Michele, supported me on the project.	



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Vipul Kashyap	Project Number S1416
Project Title Software Complexity Measurement	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals To create a software application written in Java to measure any software program code's complexity, that allows Software Developers and Software Managers to ensure that software code is simple, well documented and easy to understand by others.</p> <p>Methods/Materials Method: Application was developed using SDLC (Software Development Life Cycle) process.</p> <p>Complexity Measurements: This program defines comment complexity as: Comment Complexity = $10 - (\text{Number of Comments}) / (\text{Non trivial Code Lines})$ This program defines operator complexity as: Operator Complexity = $(\text{The maximum number of operators in a single statement}) / 2$ This program defines loop nesting level complexity as: Nesting Complexity = $\text{The maximum number of loop nesting in any loop structure}$ The overall complexity level is defined as: Average Complexity = $(\text{Comment Complexity} + \text{Operator Complexity} + \text{Nesting Complexity}) / 3$</p> <p>Materials: Computer (Dell E6410, Windows 7, Intel i5 M520 CPU, 4GB RAM 1333 mHz); Oracle Java (Version 1.6.0_23); Oracle Java JDK; NetBeans IDE (Version 6.9.1); Doxygen (Version 1.7.2); Graphviz (Version 2.26.3); MathJax (Version 2.0)</p> <p>Results After the project was completed some sample open source programs from sourceforge.net were downloaded and tested along with some self created test cases. The results provided accurate complexity measurements that could be useful to optimize software code.</p> <p>Conclusions/Discussion The project was completed successfully and with the desired results. Software Developers can now use this application and write minimally complex software code, saving thousands of dollars and hundreds of hours in maintenance for software companies.</p>	
Summary Statement This project created a new application that can measure software complexity based on defined complexity metric.	
Help Received Guidance received from Dr. Sikkema; Feedback received engineers from Cisco, Abbott Labs, and Cognizant.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Damon C. Kawamoto	Project Number S1417
Project Title Population Modeling of the Sacramento Salmon: Combining Compositional Data with Traditional Abundance Estimates	
Abstract Objectives/Goals Many challenges are faced when forecasting the abundance of salmon for the ocean-fishing harvest. I compared two methods to estimate the abundance of Sacramento River Fall-run Chinook salmon. The standard method, the jack count method, uses the number of jacks (fish that have matured one to two years early) returning to freshwater to predict the Sacramento salmon abundance for the following year. I investigated an alternative method in which the Sacramento abundance is derived using a separate abundance estimate for Klamath Chinook and the relative occurrence (determined by genetic methods) of Sacramento and Klamath fish in the harvest. I first assessed the feasibility of this new approach using historical data from the Pacific Fishery Management Council (PFMC) including the Klamath harvest, Sacramento harvest, Sacramento Index, Klamath abundance and the Sacramento jack count. Second, I applied this method to historical fishery composition data recently obtained using genetic analysis. Methods/Materials I used the statistical modeling program R to assess the strength of the relationship between the overall abundance of Klamath and Chinook salmon and their proportional representation in ocean fisheries across different seasons and coastal locations. Results Analysis of the PFMC data revealed little relationship in the overall abundance and the harvest proportion. Conclusions/Discussion Based on an r^2 measure, the jack count method is much more accurate than the alternate methods I investigated. This suggests that fishery composition estimates from genetic data are unlikely to provide substantial improvement in abundance estimates.	
Summary Statement I investigated an alternate method to estimate the abundance of the Sacramento Chinook Salmon, utilizing a separate abundance estimate for Klamath Chinook and the relative occurrence of Sacramento and Klamath fish in the harvest.	
Help Received Eric C. Anderson helped teach me how to use R, NOAA provided data needed for my project, and my advisor led me in the correct direction.	



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Meredith P. Lehmann	Project Number S1418
Project Title Stochastic Epidemic Simulations on the US Transportation Network	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals How do epidemics spread on the US transportation network? What is the role of air travel vs. long distance auto travel? Is the nature of epidemic propagation on the network such that epidemics can be mitigated with policy interventions? A large literature has simulated the spread of epidemics on the global aviation network and has concluded that large hub airports are preferred epidemic pathways at the global, national, and regional levels, simulations that involve randomness in both the infection and recovery processes within and travel across population units. This finding implies that epidemiologists can use these models to predict the spread of epidemics and to design policy interventions to mitigate it. My research is a stochastic simulation study of epidemic propagation on a model of the US transportation network that describes air and long distance auto personal and business trips to and from all 3076 counties in the continental US. Long distance auto travel has not been considered in the literature to date and auto trips might prove to be important since they are five times more numerous than airplane trips.</p> <p>Methods/Materials Each county was modeled as a homogeneous population unit described by a Susceptible-Infected-Removed model. For each day and county, random changes in the number of newly infected and recovered were simulated along with the random net travel of infecteds to and from all 3076 counties. A simulated epidemic ran until infection died out in each county. I ran 50 simulations each on the air and auto travel networks alone and 50 more with both modes operating. I also ran simulations as robustness checks using other parameter values.</p> <p>Results My simulations suggest a much diminished role for air travel and that auto travel, a previously ignored transportation mode, drives most of epidemic propagation in this model. Exponential infected growth results primarily from the local county infection process after initial seeding by infected travelers, all of which occurs in the first 7-10 weeks, who have little subsequent impact.</p> <p>Conclusions/Discussion Vaccine development must be extremely rapid or travel restrictions must be implausibly effective if this trip-based network architecture is a more accurate model than the flight based ones in the literature because, once the county infection level has reached 0.02%, an epidemic has taken root and is probably unstoppable.</p>	
Summary Statement My project shows that epidemics largely spread by auto travel in the US and that epidemics are probably unstoppable once they have taken root.	
Help Received I discussed my project with my parents and they read several paper drafts. I wrote all code in Mathematica and ran it from home on the Triton supercomputer at the San Diego Supercomputer Center, who provided 3500 hours of cpu time. Wolfram Research provided a copy of network Mathematica.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Kenny Lei	Project Number S1419
Project Title Quadrocopter Aerial Monocular Vision for Improved Autonomous Robot Navigation	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Conventional ground robot navigation and path finding is often inefficient and time-consuming, especially in a maze-like environment. Aerial monocular vision, however, provides a novel perspective in finding the path finding for robot navigation. Aerial monocular vision in combination with ground robot was compared to solely ground robot navigation for operational time.</p> <p>Methods/Materials A ground robotics platform was based off an iRobot Create and laptop. Aerial vision was achieved through the Parrot AR.Drone quadrocopter with a built-in camera. A laptop was connected to the camera feed of the quadrocopter via socket connections to its wireless network. Java programming language was used for both quadrocopter control and image processing. The quadrocopter was initiated and hovered above the robot and maze environment. Images acquired were initially processed to classify regions as either obstacle or traversable area. Start and end point regions were then classified within the image. A breadth first search (BFS) algorithm was employed to determine the shortest navigational path that avoids obstacles. When a traversable path between the detected start and end points is found, the ground robot is sent movement vector commands to navigate around the obstacles.</p> <p>Results After a series of trial runs, the novel navigation yielded an average run time of 38.45 seconds while the conventional navigation resulted in an average run time of 140.57 seconds. The addition of aerial vision from the quadrocopter resulted in a 72.6 percent improvement in operation time for the ground robot.</p> <p>Conclusions/Discussion These findings demonstrate rich data provided from aerial monocular vision significantly enhances and improves robot navigation. The increased complexity of a multi-modal robotics platform yielded improvements in navigation time.</p>	
Summary Statement Aerial monocular vision in combination with ground robot was compared against solely ground robot navigation in operational time.	
Help Received Used robotics lab at Harvey Mudd College under the mentorship of Dr. Dodds.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Jonathan F. Li	Project Number S1420
Project Title Effects of Cell Compressibility, Motility, and Contact Inhibition on the Growth of Tumor Cell Clusters	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals To study how tumor growth are affected by cell compressibility, motility and contact inhibition.</p> <p>Methods/Materials I created a mathematical model and computer simulation of tumor cells. I incorporated properties such as compressibility, motility, adhesion energies, and contact inhibitions as well as the 4 major phases of the cell cycle to simulate growth and mitosis. After 30 simulation runs, I input the data to MATLAB for statistical analysis. Lastly, I compared results obtained from in vitro experimentation to the simulation data.</p> <p>Results I discovered an inverse relationship between the motility and compressibility parameters. I used this result to calibrate the rest of my simulations. I discovered that tumor cell clusters that are not compressible grow faster than those that are. In addition, compression plots show that cells in the center of the cluster tend to be the most compressed in the cluster. I also test the motility parameter and find that motile cell lines have fast growth rates. Finally, comparisons with in vitro experimentations have shown that my model is accurate.</p> <p>Conclusions/Discussion I found that motility is an attribute that can make cells more fit if motile cells do not experience significant compression, compared with their less motile counterparts. Motility plays an especially important role when contact inhibition is a factor. Contact inhibition acts as a penalty for clustered cells since growth is halted when cells are surrounded by other cells. However, I have also found that increasing cell motility can induce changes in cell volume as it is easier for smaller cells to move. My research identified the interplay between cell motility and stiffness via the system energy and provided a means for compensating for this effect. This problem is an artifact of all simulations using the Cellular Potts Model and the procedure we discovered to compensate for the interplay may be used for all models.</p> <p>These results also call into question the effectiveness of cancer therapies that involve high cell death rates such as chemotherapy. The mass die-off of cells places a selective pressure on more motile cells. Each round of die-off increased the proportion of motile cells. Similarly, high cell death rate therapies could select for cells with a particularly low response to contact inhibition, leading ultimately to more invasive cells.</p>	
Summary Statement I developed a mathematical model and computer simulation to study the effects of cell compressibility, motility and contact inhibition on the growth of tumor cell clusters.	
Help Received I wrote the draft report and my mentor provided valuable comments for me to revise the paper in its final form.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Christian Liu	Project Number S1421
Project Title Effect of Gap Penalty on Global Alignment Accuracy	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The Needleman-Wunsch global alignment algorithm uses a scoring system consisting of a match, mismatch, and gap penalty to compare two large sequences of data. This research aims to examine effects of the gap penalty on sequence alignments when the match score is 1 and the mismatch score is 0.</p> <p>Methods/Materials Sequence-modifying, global alignment, and alignment scoring programs were written in Java and used to compare alignments. The sequence-modifying program controlled how similar sequences were by altering 50, 200, 350, 500, and 600 characters in an original 626-character long sequence. These altered sequences were then globally aligned to the original DNA sequence using 0, -1, -2, -3, and -5 as gap penalties. Finally the scoring program judged the accuracy of the alignment.</p> <p>Results At a gap penalty of 0, there were gaps throughout the alignments and alignment scores were low, but at a gap penalty of -5, the algorithm avoided gaps for all alignments. The ideal gap penalty was -2.</p> <p>Conclusions/Discussion At very small gap penalties, scores are low because the algorithm doesn't penalize for using many gaps. However, the algorithm quickly stops aligning as penalties grow, so the most effective penalties are small values still large enough to discourage unnecessary gaps.</p>	
Summary Statement Five gap penalties were used to align sequences against an original and model the effect of gap penalties on the accuracy of the global alignment algorithm.	
Help Received My advisor Dr. James Li reviewed the programs I wrote and helped me correct errors.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Galina H. Meyer	Project Number S1422
Project Title The Application of Bayesian Networks for Speech Classification	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals Computers that can process the large amounts of data statistics requires have made available unprecedented means of understanding the word, resulting in radical advancements in everything from quantum physics (modeling the behavior of subatomic particles using statistics) to environmental science (modeling changes in weather) to stock market analysis (modeling micro- and macro-economy trends). One of the final frontiers of science is understanding the human mind, and how it can communicate ideas. In this project, I wanted to explore applying a specific field of probability, called Bayesian networks, to identify what abstract idea some body of text is conveying-- here, whether a politician is advocating Democratic or Republican ideas.</p> <p>Methods/Materials I typed my program in LISP (dialect: Racket), using the IDE called DrRacket. My corpus of speeches, called Corps, came from Foundation Bruno Kessler and was generously supplied to me by Mr. Guerini and Mr. Strapparava. I mostly used Artificial Intelligence: A Modern Approach by Stuart J. Russell and Peter Norvig for my research. My paper was formatted in LaTeX, with the IDE TeXworks. I created a Bayesian network, with values determined from training the computer with the corpus, and used Bayes Theorem to derive the probability that a speech was from a certain party given the words that were in it.</p> <p>Results The program works, returning surprisingly accurate results. For extremist politicians such as Huckabee, it returned 99.8% accuracy, and for more moderate politicians, such as Ronald Reagan right after he changed to the Republican party, it returned 75.4% Republican. These trends continue fairly accurately over a wide variety of politicians, even those from overseas.</p> <p>Conclusions/Discussion The success of this program signals that computers really can be used to extract abstract ideas from a list of words, partially understanding natural languages by observing the trends of human speech. Humans learn in a manner similar to this, by listening to years worth of conversation, so this program actually follows how humans learn a language to a certain extent. This program is not limited to a Republican versus Democrat categorization: it can also include more categories, and more abstract. One possible option is identifying a speech that promotes violence or is</p>	
Summary Statement I categorize political speeches with a Bayesian network.	
Help Received Mr. Anderson gave me the AI: Modern Approach book	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Ani Pochiraju; Rohan Rastogi	Project Number S1423
Project Title Bimodal Ultrasound and Optical Approach for Breast Cancer Imaging without Coregistration	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The aim of this investigation is to test the potential and efficiency of a bimodal ultrasound and optics probe without coregistration in locating a breast cancer lesion in standard breast tissue. Traditionally, there exist many separate techniques that achieve this said task: ultrasound scanning, diffuse optical spectroscopic imaging (DOSI), MRI, CAT scans, etc. Our novel approach combines the strengths of the first two conventional methods of lesion detection into one multi-capability probe. It was hypothesized that we would be able to locate a cancerous lesion in breast tissue, and that the results from our bimodal optical and ultrasound (DOSI/UZ) approach would follow expected biomarker trends.</p> <p>Methods/Materials The probe was designed and manufactured with black delrin plastic, steel 5 cm. screws, steel 0.32 cm. set screws, silicone rubber, a clinical ultrasound probe and machine, and DOSI optical fiber imaging technology. Measurements of biomarker concentration and tumor size, position, and depth were then obtained from a breast cancer patient with the help of a trained clinician using our bimodal probe. Consequently, the capabilities of the probe were analyzed using MATLAB software by comparing the biomarker levels of lipid, water (H₂O), tissue optical index (TOI), hemoglobin (HbO₂), and deoxyhemoglobin (HHb) to the theoretical trends found in our research. These biomarkers provide functional parameters related to metabolism, angiogenesis, and cell/matrix density of the breast tumor.</p> <p>Results As tumor depth increased and the amount of light from the optical instrument interacting with the tumor decreased, we expected TOI to decrease, lipid levels to increase, hemoglobin levels to decrease, and water levels to decrease. Our results followed these expected trends for a comparison between biomarker concentration in the tumor and the depth of the tumor.</p> <p>Conclusions/Discussion Our results show that our bimodal ultrasound and optics probe was successful in locating the breast tumor and that the data obtained from our bimodal device was conclusive. Further patient tests will confirm the overall superiority of our technique to conventional ones.</p>	
Summary Statement We tested the potential and efficiency of a new bimodal optical and ultrasound multi-capability probe to image and locate cancerous tumors in breast tissue.	
Help Received Used lab equipment at the Beckman Laser Institute at the University of California, Irvine under the supervision of Dr. Cerussi; basic lab and technological assistance provided by Dr. Darren Roblyer, Shanshan Xu, and Zheng Zhu	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Charlie A. Reid	Project Number S1424
Project Title Simulating Incompressible Fluid Flow with the Navier-Stokes Equation	
Abstract Objectives/Goals The objective was to determine how accurate my fluid simulator was. Methods/Materials The simulation method was a staggered grid finite difference approximation of the two dimensional incompressible Navier Stokes Equations. To test the method's accuracy I simulated fluid flow around circles of varying size. I programmed the whole program in java from scratch. Results The simulation produced results that looked like water in general but did not agree with past experiments numerically. The frequency did decrease with cylinder size as expected but it was almost exactly ten times more than it should be. Also, the frequency decreased about half as much, proportionally, as expected for the higher diameters. All the simulations converged to a stable frequency after less than ten vortices. Conclusions/Discussion This experiment showed definitively that my simulator was too inaccurate for anything but possibly computer graphics. However, the cause of the inaccuracy is still uncertain. A question for future exploration is whether the error approaches zero as the grid resolution approaches infinity and the time step approaches zero. If so, then the approximation method needs to be improved and possibly greater computing power must be used. If not, then there is probably a bug in the program or something fundamentally wrong with the math.	
Summary Statement My project was an attempt to make an incompressible two-dimensional fluid simulator that agreed with the literature.	
Help Received I had some discussions with my science teacher on related subjects such as boundary layers on hypersonic aircraft.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Yousuf Soliman	Project Number S1425
Project Title Cube It: Creating a 3D Rubik's Cube Simulator in C++ and OpenGL	
Abstract Objectives/Goals The objective of this project was to program a 3D Rubik's Cube simulator in C++ and OpenGL. The application should be able to scramble and solve the Rubik's Cube. Along with that, the application should allow the user to put in their own position and have the application generate a moves list needed to solve it. Methods/Materials The program I wrote has a graphical user interface in which the user can view the Rubik's Cube in 3D or as a 2D net. The application can scramble and solve the Rubik's Cube. The computer scrambles the cube by choosing a random face, executing a move on that face, then moving on to another face; the application does this 30 times per scramble. The application also has the ability to have the user input their own position. For every solve the application will generate a moves list needed to solve the cube. Results After the project was completed, many runs were taken. These runs helped show truly how the application should be rated. One hundred thousand runs were executed to see how many moves on average it would take to solve the cube. On average, it was able to solve it in about eighty-one moves. The minimum number of moves was thirty-nine, and the maximum was one hundred and six moves. The application was very successful and was able to accomplish its task. Conclusions/Discussion This application has been working flawlessly 100% of the time. The application was able to have the user input their position and the program will solve it. In addition the application can scramble and solve it on its own. The user can also view the cube in 3D or as a 2D net. If the user is viewing it in a 2D map they will have the option to edit the position of the cube. After the application is told to solve the cube it will generate a moves list needed to solve it. Although the application solves the Rubik's Cube fairly quickly, more complex algorithms can result in faster speeds and less moves. The optimal solution would be God's Algorithm, this solves the cube in the minimum amount of moves needed.	
Summary Statement For this project I programmed a 3D Rubik's Cube simulator in C++ and OpenGL.	
Help Received My computer science teacher, Mrs. Najwan, for honing my programming skills and teaching me object oriented programming; My mother for helping me with my board.	



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Rahul Sridhar	Project Number S1426
Project Title Multi-document Summarization using Spectral Clustering	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals It is not uncommon in today's electronically connected world to get information about the same subject from a variety of sources. Search engines like Google have made this possible with a mouse click and as a result human beings are inundated with information. The challenge is to combine these "results" into a concise summary. Can this summary be automatically generated based on quantitative scores with no qualitative judgment? In other words, can we write a program that will create a concise, effective, and coherent summarization of multiple articles on the same subject?</p> <p>Methods/Materials My method uses the technique of spectral clustering to summarize multiple documents. Clustering techniques are widely used for data analysis and spectral clustering often outperforms traditional clustering algorithms such as k-means. Given a set of documents, the program will first create a similarity graph, with vertices representing the sentences in the documents and weighted edges between vertices to represent sentence similarity. Next, the graph will be divided into a certain number of clusters, where each cluster represents a group of sentences that are similar to each other. A representative sentence is then chosen from each cluster. These sentences are then ordered to create a summary. For my experiments, I chose news articles and results of search-engine queries as multi-documents.</p> <p>Results The proposed method is fast and effective for documents containing news articles and reviews. In addition, the results satisfy two key properties: (i) summary does not contain redundant information; (ii) sentences conveying little or no information are not included in the summary.</p> <p>Conclusions/Discussion In this reasearch, I implemented a method to generate a concise and accurate summary of multiple documents on a common subject. This research would have applications in many varied fields, for example in summarizing news articles or search results of specific topics. For future work, I would like to extend this technique to implement an "aggregator" that can collect multiple related news articles from a set of sources that feeds into the "summarizer".</p>	
Summary Statement Given multiple documents on a related subject, automatically create a summary that is both coherent and accurate.	
Help Received My Mother helped me with proof reading my slides.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Kevin T. Tang	Project Number S1427
Project Title Characterizing the Optimal Flight Time of a UAV Helicopter, an Odyssey through Numerical Simulations	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This project deals with optimizing the flight operation of a UAV (Unmanned Aerial Vehicle) helicopter. I intended to characterize the optimized minimum flight time to hover between two end points. My goal was to develop a relationship for estimating the minimum flight time without using the complicated optimization method.</p> <p>Methods/Materials The tools I used for this project include an optimization program and a UAV helicopter model. MATLAB was used in data analyses. The optimization program applied to the nonlinear helicopter model resulted in a solution optimized for minimum flight time for two given end points. In this process, initial guesses of controls and state variables were fine tuned in search for an optimization solution. Solutions were obtained for some sample end points with various linear distances. I analyzed these results and established an empirical relationship between the minimum flight time and the distance using least square linear regression method. I validated this relationship using independent samples from the full optimization process. I further examined the effects of the number of nodes in the optimization process on my results.</p> <p>Results I found that the optimal flight time is best described as a linear function of the distance even though the optimized flight trajectory is rather complicated. My regression results show very small bias and standard error when compared to the full optimization solutions. The function is also applicable over an extended range of distances. The number of nodes may change the flight trajectories, but it does not significantly affect the total flight time.</p> <p>Conclusions/Discussion I obtained an empirical function between the minimum flight time and the linear distance between the two end points for a UAV helicopter. This relationship was found to be accurate and robust.</p>	
Summary Statement This project is about defining the optimal operation envelop of a UAV helicopter without using the computationally intensive optimization process.	
Help Received Professor Wei Kang provided me with the optimization program (snOpt) and the nonlinear helicopter model developed by the National University of Singapore.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Pauline Truong	Project Number S1428
Project Title Development of a Mashup That Transforms a Smartphone into an Assisted Vision Aid	
Abstract Objectives/Goals The purpose of this project is to (1) develop a smartphone-based mashup that transforms a smartphone into an assisted vision aid and (2) measure its effectiveness. Methods/Materials A prototype was developed and tested using an Android smartphone with gesture recognition, image capture, orientation sensors, text-to-speech software and an image recognition service. Test subjects were taken to an outdoor test environment, blindfolded, guided to a random starting location and orientation, and instructed to identify the location of a given object. Four trials runs using the prototype and four control runs were performed. Success was measured in terms of whether the test subjects could identify the location of the target object within a given time limit of 5 minutes. Results 75% of trial runs with the prototype system were successful while 100% of the control runs failed. The average time it took for test subjects using the prototype to find the target object was 3.02 minutes. Since all control runs failed, the average time of the control runs was the 5.00 minute maximum test time. Conclusions/Discussion This experiment has shown that the sensors available on commonly available smartphones can be integrated with new image recognition services to assist the blind and visually impaired. The degree of assistance (75% success rate) is significant in light of the fact that the control group failed 100% of the time given the same objective of finding a large object in a relatively uncluttered outdoor environment within 5 minutes.	
Summary Statement A smartphone-based mashup was developed and tested to serve as an assisted vision aid for the blind and visually impaired.	
Help Received My family donated a smartphone for my use. I also received guidance from my teacher, developer blogs and message boards, and open source projects.	



**CALIFORNIA STATE SCIENCE FAIR
2011 PROJECT SUMMARY**

Name(s) Alanna K. Williams	Project Number S1429
Project Title An Application of Geographic Profiling to Graffiti Crimes	
Abstract Objectives/Goals The purpose of my project was to see if it is possible to successfully use an already established formula(Rossmo's formula), which is used to predict the residences of serial murderers, rapists and arsonists based on their past crimes, for graffiti crimes. Methods/Materials This process involved taking a set of known graffiti locations and matching them with four known vandalswhose residences were approximately known. Using Matlab#, I was able to write programs that allowed me to perform various regressions to solve for empirically determined constants and exponents in Rossmo's formula. Theoretically, these values could then be used to try and predict the locations/residences of unidentified graffiti vandals. Data on graffiti locations and vandal residences were acquired by a combination of interviews with the Santa County Sheriff's Department, the Santa Cruz city planning department; there was also some data simulation required Results After all empirically determined values were solved for to most accurately fit the crimes of all four vandals, the algorithm in general did not produce probability contours that consistently overlapped with the actual offender residences. Conclusions/Discussion Overall, the algorithm has very minimal predictability value for graffiti crimes. Either a different equation needs to be utilized or mathematical modeling is not the most efficient means of locating offenders.	
Summary Statement My project is to examine the usefulness of an algorithm, which is used to catch seriel criminals, when applied to graffiti crimes.	
Help Received My parents, Quentin Williams and Elise Knittle,let me use the MatLab Software that they have in their geology lab at UCSC. Sheriff Robin Mitchell gave me information on specific vandals and graffiti patterns in general. J. Guevera provided me with a database of graffiti crime addresses in Santa Cruz City.	



CALIFORNIA STATE SCIENCE FAIR 2011 PROJECT SUMMARY

Name(s) Mike Wu; Stephen Yu	Project Number S1430
Project Title Position and Vector Detection of Blind Spot Vehicle with Horn-Schunck Optical Flow	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals The main goal is to try to create a cost-effective way to find information about a vehicle present in the blind spot. Our three calculation goals are the object's position, velocity and acceleration in the blind spot. We based our coding off of the Horn-Schunck Optical Flow algorithm to make statistical analysis based on the vector plot of motion.</p> <p>Methods/Materials The materials used were limited to a minimal for conservation: a video camera and a coding system.. The overall Optical Flow method was adjusted with a convulsion approach, meaning that a grouping of pixels is compared within two subsequent frames. Since each specific pixel has a unique luminescence, Optical Flow is able to trace the pixel onto the next frame and draw a vector arrow to represent the movement. We then split the vectors into car and background vectors. To detect the position, we created a box to surround the object of motion and then trace it throughout the video. The center of the box is found by averaging the (x,y) coordinate of all the car vectors while the size of the box is directly proportional to the ratio of car to background vectors. Then a standard deviation filter is used to rule out any outlier vectors X standard deviations away from the initial center, till an accurate new center is found. Further improvements like a threshold for the box movement were made to prevent miscalculation; a beeping system was created to alert the viewer when a box was formed. Velocity could be estimated from the movement of the center of the box and acceleration is easily calculated through a derivative.</p> <p>Results We coded the output rose plot and quiver plot to play in a movie format. These video trials will be available for viewing during the fair. Since the project was more conceptual, no actual numbers are presentable.</p> <p>Conclusions/Discussion In the trials that we ran, the program was able to accurately detect the vehicle and follow it throughout the frames. It was also able to give a decent estimation of relative velocity. The main goal from now is to get the program to run in real time. This way, we can see exactly whether not the code would function properly in real life implementation.</p>	
Summary Statement To detect the presence of motion in the blind spot by using a video camera and computer coding software in order to alert the driver of the position, velocity and acceleration of any changing object in the blind spot.	
Help Received Brendan Morris helped introduce Matlab coding to us and teach us coding strategies .	