Culminating Project

Exercise and Self-Esteem in High School Students

MDM4U

By
Mary Agate

Culminating Project

MDM4U
By
Mary Agate

Does the frequency of exercise by high school students influence how they feel about their lives or themselves?

Section 1- searching for an area of interest and locating data

I began this project by exploring a number of areas that interested me. These areas were what universities I might apply to, the use of alcohol/drugs by high school students, and the effect of sports participation on teen's mental and/or physical health. I decided to investigate the sports and health information as it interested me the most.

Mind Map
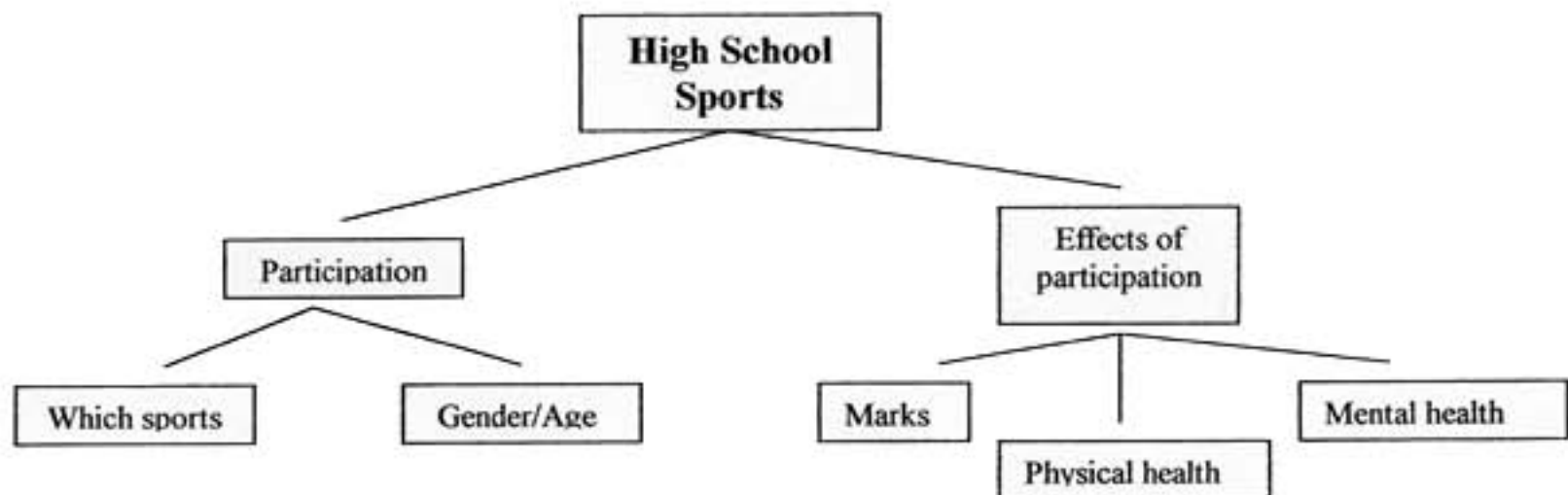
I first wondered if high school students who participated in sports were more successful in school academically. I began my search in Google, looking for *high school participation, Canada*. The first site I looked at was:

www.pch.gc.ca/progs/sc/pobs/status_e.cfm

At this site, I was introduced to a 1994 Stats Canada research paper on sports participation in Canada. There was information about gender, age, income and different sports for Canadians over the age of 15. I felt that this would be helpful at various stages of the project.
Another site with interesting information was:

http://www.coach.ca/getmoving/fact2_e.htm

Although the information was interesting here, there was no data that was available to use mathematically.

A large report that I finally saved to a CD was a 1998 report of one hundred and forty eight pages. Here I found statistics about families and sports, spending on sports, and health benefits of sport. This was found at:

http://canada.gc.ca/publications/publication_e.html

The most helpful data was found in E-Stat.

http://estat.statcan.ca/content/english/overview/edu-o.shtml

At E-Stat, I choose the English path and accepted the agreement and typed in the password. From the Table of Contents, I looked at Health and Education and found a number of articles and documents that would be useful. One of the documents in the Health section was the *National Longitudinal Survey of Children and Youth: Participation in Activities, 1989/99.* There were a number of graphs here that gave me ideas of the kind of data I might be able to analyze here.
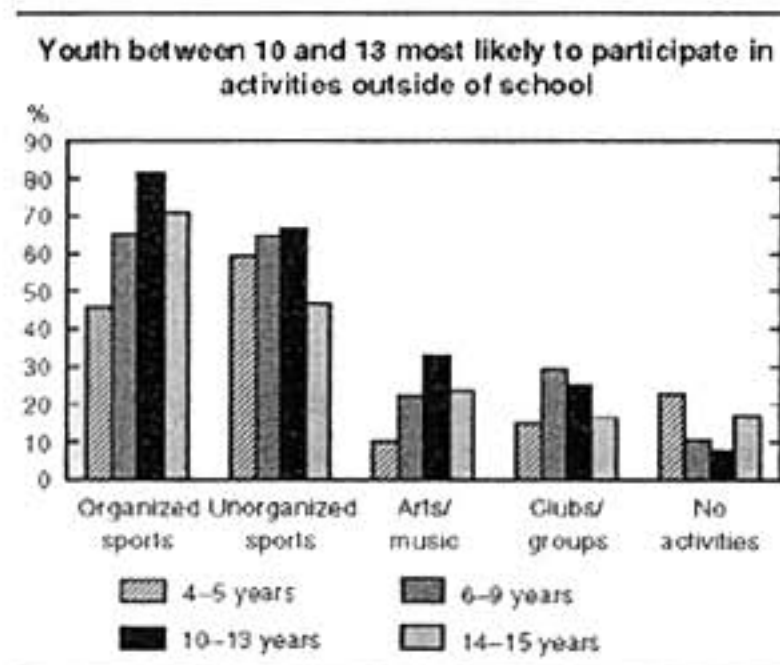
By this stage I had narrowed my hypothesis down to these two possibilities.

Hypothesises:

1. Involvement in team sports helps students live a healthier life, both physically and mentally.
2. Participation in sports has a positive influence on the grades and self esteem of high school students.

Section 2 – Applying various mathematical techniques in the analysis and exploration of the data.

Section 2a – Analyzing data involving one variable

Problem: In this section of the project I used two websites for information.
http://estat_statcan.ca
http://canada.gc.ca/publications./publication_e.html

I had some difficulty finding raw data here and ended up using data from the articles on my research topic. The first graph showed the frequency of exercise outside of school by 15-year-old Canadian males and females outside of school.

Frequency of exercise outside of school by students in selected countries

1998 - 15 years; Canada
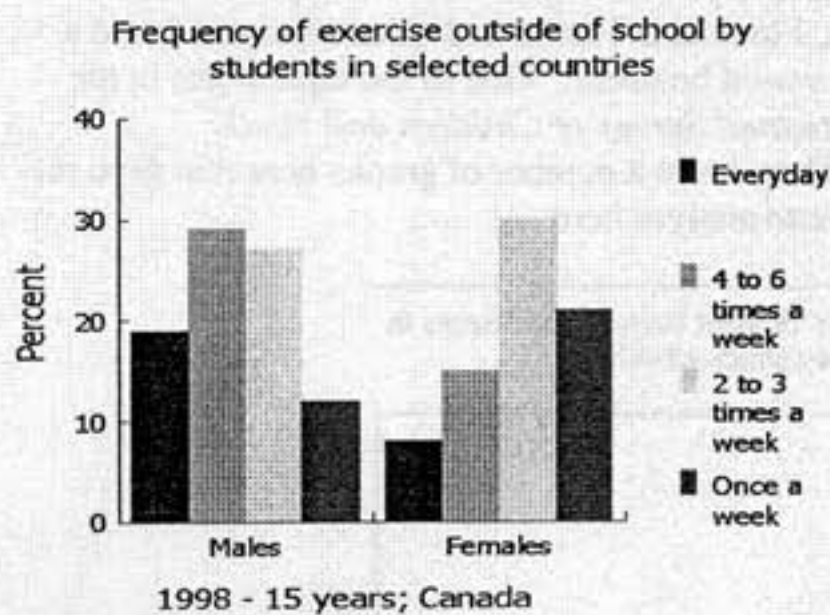
From this graph I could determine that 15-year-old males most often exercised at the 4-6 times/ week rate and that the females most often exercised at the 2-3-times/week rates. From this graph I was able to determine the mode for 15-year-old males (4-6) and 15-year-old females (2-3).

The next data I used was from the 148-page publication from the Government Publications site. I made a graph to display how active teenagers perceive the benefits of sports participation. I was unsure how to continue with the data I had as it was all in graph and chart form, no raw numbers to work with. I would however find these raw data numbers at a later stage in the project work.

Rank of Importance of Participating in Sports

Very Important (%)
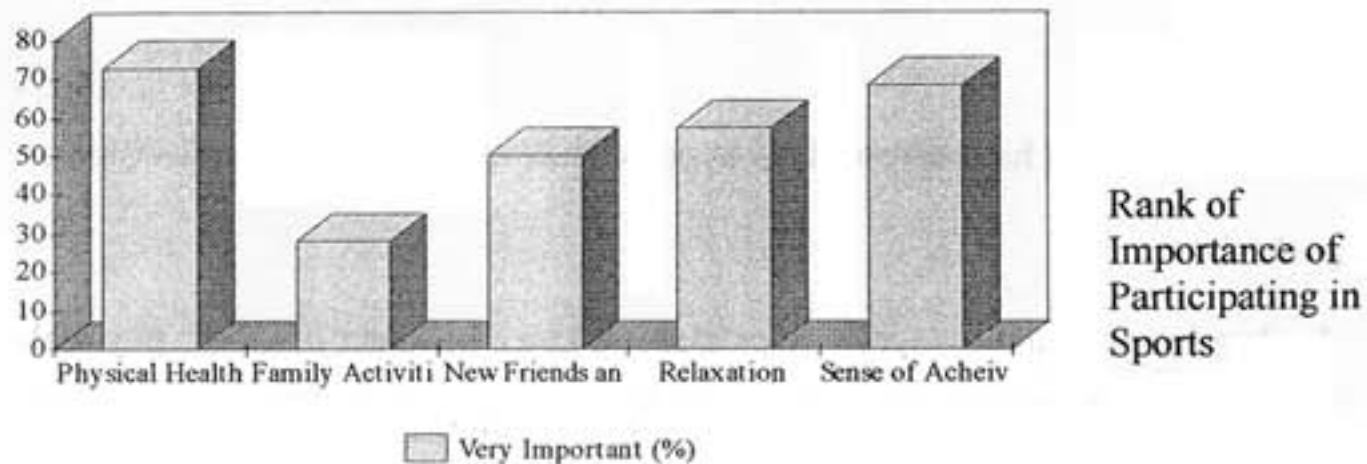
I also found data in chart form here that illustrated the difference in participation levels between boys and girls from age 4 to age 15. This was interesting as it showed that girls at ages 4-5 are more active in sports than boys, but that they then trail at all ages after that. After girls hit the age of 13, their sports participation drops markedly. I wondered if this was also a time when self-esteem dropped in girls also, and if the two things were related. The final data I analyzed in this section was from E-Stat and it was a graph showing how sports participation related to self-esteem factors. It showed a relationship between these two things.

Organized Sports Participation Related to Factors such as Self Esteem and Socialization for Youth Age 12-15
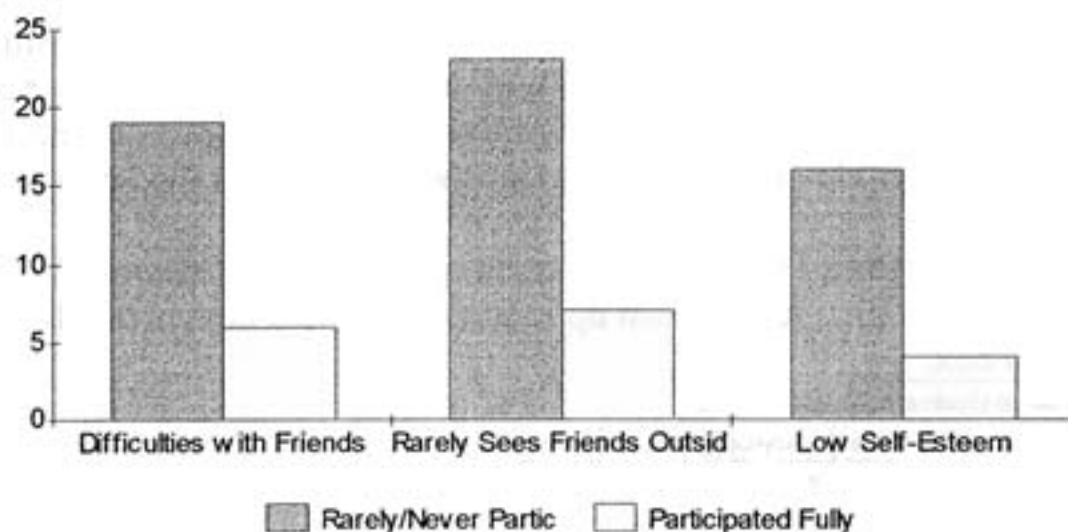
Conclusion: The data here showed that males tend to exercise more frequently than females at the age of 15 in Canada. I also found that the most commonly perceived benefit of sports participation was physical health, followed closely by a sense of achievement. There appears to be a relationship suggested between sports participation and self-esteem factors in Canadian youth and I hoped to find more data to manipulate as the project continued. I began to feel that I was on the right track but I was frustrated by not finding numbers to work with.

Section 2b – Analyzing data with two variables

Problem: At this stage of the project I refined my hypothesis: "Participation in high school sports has a positive influence on the grades and self esteem of high school students."

I now refined the question to: "Does the frequency of weekly exercise in high school students influence how they feel about themselves, or their lives?"

I went back to E-Stat and found data that included a number of different countries, the frequency of exercise in teens, and questions about positive outlooks on life. I decided to narrow this data down to percentage of 15-year-old females who exercise at a frequency of "4-6 times per week", "feel very happy about life", and "never feel lonely". I eliminated countries with incomplete or no data from my study (Scotland and Belgium). The correlation coefficient of these two scatter plots showed a very weak positive linear relationship. I noticed two outliers in these graphs and since least-square regression lines are very sensitive to outliers, I thought I would investigate the outliers. I removed two cases from the collection (Slovakia and Spain), as they were either incomplete or perhaps biased.

I then retrieved the data from 1998 as a table and organized it in a Microsoft spreadsheet. This eliminates the need for a lot of clean up in Fathom. I then copied the data into a Fathom collection box.

Exercise and Positive Outlooks (Spain and Slovakia removed)				
	Country	Fourtosixtimeswk	Veryhappy	Neverlonely
1	Austria	15	38	22
2	Canada	15	26	19
3	Czech Repu...	19	22	8
4	Estonia	16	13	13
5	Finland	22	27	23
6	France	10	32	23
7	Germany	16	16	24
8	Greece	11	20	11
9	Greenland	12	19	16
10	Hungary	6	14	12
11	Ireland	15	25	22
12	Israel	11	15	16
13	Latvia	6	10	7
14	Lithuania	8	11	7
15	Northern Irel...	16	33	26
16	Norway	16	38	36
17	Poland	10	20	12
18	Portugal	9	16	10
19	Russia	10	10	18
20	Sweden	14	34	35
21	Switzerland	15	37	28
22	United States	12	22	20
23	Wales	16	20	29
24	Denmark	15	26	47

I used the data to create scatter plots for each of the following.
Exercise 4-6 times per week vs. Feel very happy about life AND
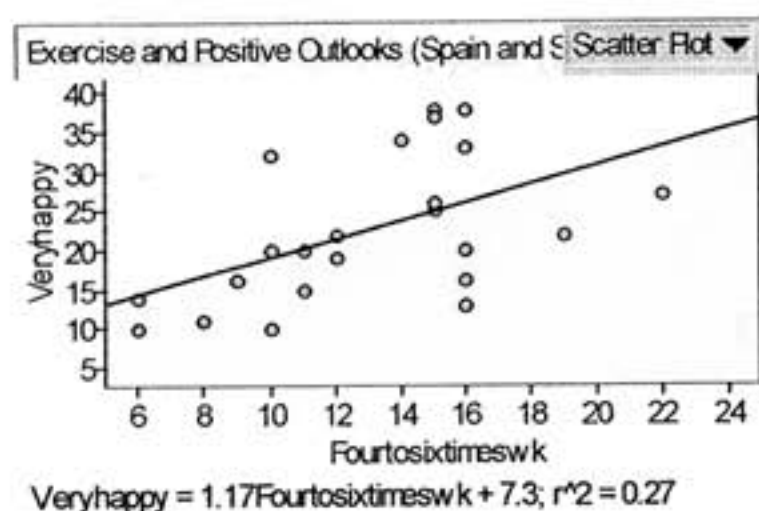Exercise 4-6 times per week vs. Never feel lonely.

Veryhappy = 1.17Fourtosixtimeswk + 7.3; r^2 = 0.27

Figure 1

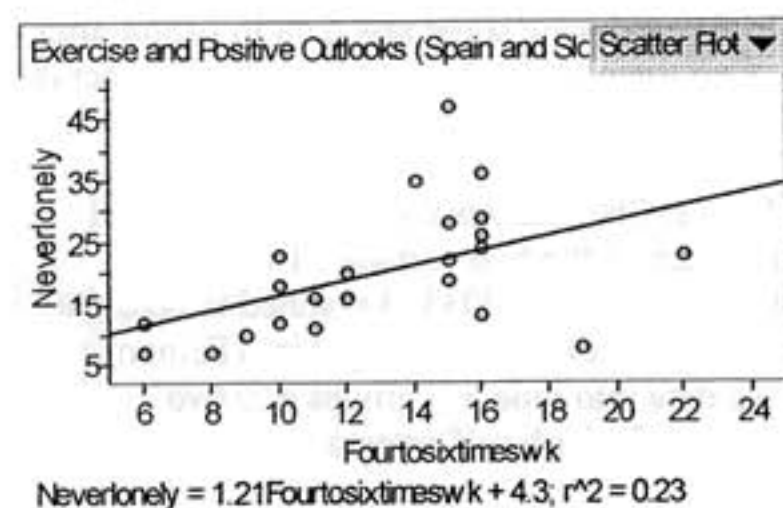Neverlonely = 1.21Fourtosixtimeswk + 4.3; r^2 = 0.23

Figure 2

Analysis:
Using Fathom, I did a linear regression for each scatter plot and a line of best fit appeared
with the following values.

Graph	Fourtosixtimeswk(x) vs. Veryhappy(y)	Fourtosixtimeswk(x) vs. Neverlonely(y)
Equation of line of best fit	$y = 1.17x + 7.3$	$y = 1.21x + 4.3$
r^2	$r^2 = .27$	$r^2 = .23$
Correlation coefficient	$r = .52$	$r = .48$

To see how well the line fits each graph a correlation coefficient was calculated from the
r^2 value given by Fathom.

Both values indicate that there is a moderate positive correlation.

Conclusion:

There exists a moderate positive linear relationship between 15-year-old females who exercise 4-6 times per week and the statements "I feel very happy about life" and "I never feel lonely".
The equations of the line of best fit are:
Very happy = 1.17(Four to six time/wk) + 7.3
The correlation coefficient = .52

Never lonely = 1.21(Four to six time/wk) + 4.3
The correlation coefficient = .48

Section 2c – Probability distributions

Problem: I want to explore with Fathom the probability of having high self-esteem at different ages and compare it to the probability of being physically active at different ages. At this stage of the project I was finally able to locate some raw data in E-Stat after spending a lot of time looking through files. One thing I have learned from this project is that researching requires a lot of time reading and sifting through realms of information.

Plan: I planned to collect the data I wanted from E-Stat and work with it in Fathom. I had to use data from 1994 in order to have numbers on the two variables I wanted to research. The tables used from E-Stat are 104-0019 and 104-0033. I planned to take the data from E-Stat and put it into a Microsoft spreadsheet before putting it into Fathom, as it is easier to clean up the data. I organized the data into nine age groups and two collection groups, one for physical activity and one for high self-esteem.

Relative Frequency Graph of Physical Activity by Age Groups

Probability work with Physical Activity

|	Age	MFPA	tpSurve...	rfMFPA	FPA	tpFSurv...	rfFPA	MPA	tpMSur...	rfMPA
1	12-14	505596	1326127	0.381258	189529	620386	0.305502	316067	705741	0.447851
2	15-19	718468	2046278	0.35111	249076	989002	0.251846	469391	1057276	0.443963
3	20-24	362197	1739550	0.208213	152164	902132	0.168672	210033	837418	0.25081
4	25-34	810956	4755495	0.17053	352572	2417732	0.145828	458384	2337763	0.196078
5	35-44	692405	4864334	0.142343	331331	2390243	0.138618	361074	2474091	0.145842
6	45-54	507606	3504062	0.144862	232071	1687014	0.137563	275535	1817048	0.151639
7	55-64	404157	2461802	0.164171	184807	1309418	0.141137	219349	1152384	0.190344
8	65-74	323129	2059918	0.156865	123418	1139707	0.108289	199711	920211	0.217027
9	75+	124113	1191051	0.104205	54926	712641	0.0770739	69187	478410	0.144619

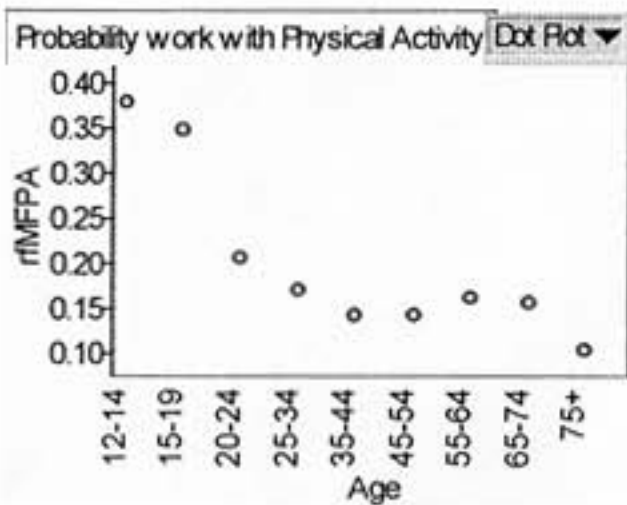Relative Frequency Graph of High Self-Esteem by Age Groups

Probability Work with Self Esteem

	Age	MFHSE	tpSurve...	rfMFHSE	FHSE	tpFSurv...	rfFHSE	MHSE	tpMSur...	rfMHSE
1	12-14	475646	1326127	0.358673	238692	620386	0.384748	236953	705741	0.335751
2	15-19	833809	2046278	0.407476	369371	989002	0.373479	464438	1057276	0.439278
3	20-24	806706	1739550	0.463744	402080	902132	0.4457	404626	837418	0.483183
4	25-34	2403301	4755495	0.505373	1219067	2417732	0.504219	1184234	2337763	0.506567
5	35-44	2564859	4864334	0.527279	1224220	2390243	0.512174	1340639	2474091	0.541871
6	45-54	1877333	3504062	0.535759	933586	1687014	0.553396	943748	1817048	0.519385
7	55-64	1200827	2461802	0.487784	650793	1309418	0.497009	550034	1152384	0.477301
8	65-74	962159	2059918	0.467086	557841	1139707	0.48946	404317	920211	0.439374
9	75+	525261	1191051	0.441006	317701	712641	0.445808	207560	478410	0.433854

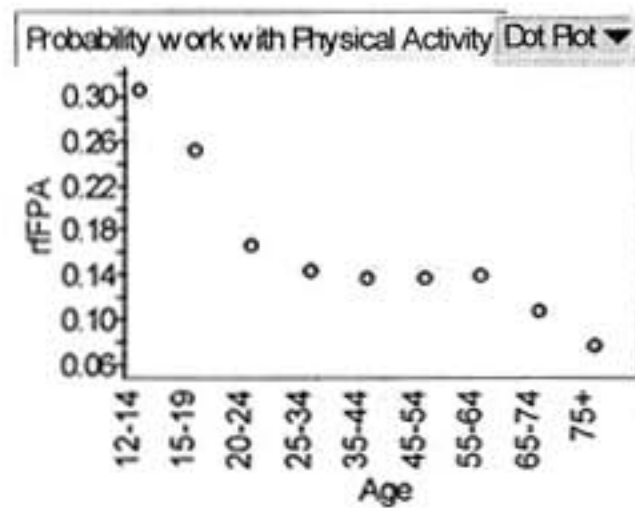I then plotted the data into separate dot plot graphs.

Graph 1: Relative Frequency of Males & Females who are physically active.
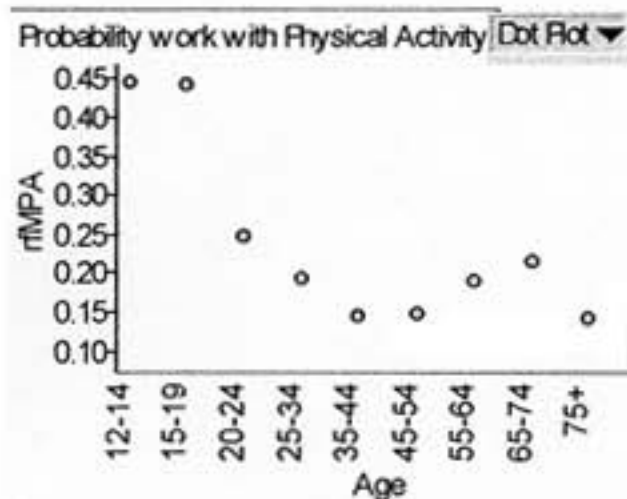
Graph 2: Relative Frequency of Males who are physically active.

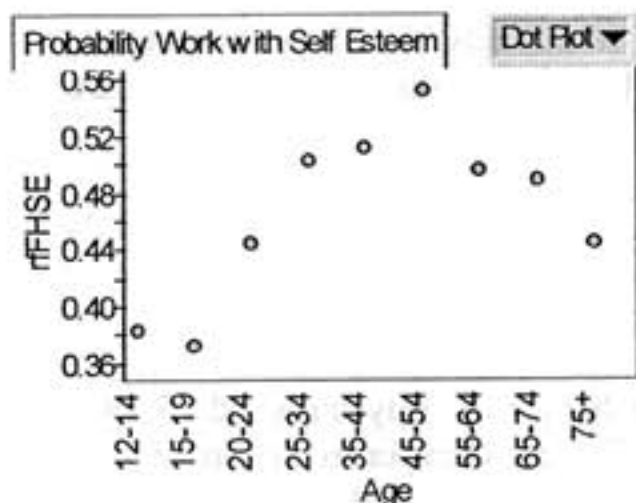Graph 3: Relative Frequency of Females who are physically active.

Graph 4: Relative Frequency of Males & Females who have high self-esteem.

Graph 5: Relative Frequency of Males who have high self-esteem.

Graph 6: Relative Frequency of Females who have high self-esteem.

Analysis: Graphs 1,2 and 3 show a downward trend in Canadians who report being physically active. There is a sharp drop in activity after age 19, and a steady drop after age 64, except in males who appear to increase activity levels after age 44 until 74. Females remain more consistently active compared to males, but are less active in the age group of 15-19, the high school years.

Graphs 4,5 & 6 show a bell curve showing that feelings of high self-esteem peaks for males around age 44 and for females around age 54. Females in the age group 15-19 have a drop in high self-esteem, while males in this high school age group increase the probability of having high self-esteem.

Conclusion: When comparing the physical activity graphs to the high self-esteem graphs, the occurrence of physical activity doesn't appear to coincide with high self-esteem. The age group that I was most interested in, the 15-19 year olds had the most interesting data in these graphs. While this age group has almost the highest physical activity level, they also have the lowest probability of having high self-esteem. I expected to see lower physical activity levels for Canadians after age 65 but did not really expect to see such a drop in self-esteem also. I also found it interesting that females peak in feelings of high self-esteem later than males and wonder what the reasons for that could be. I think however, with respect to the high school age group of Canadians, it is important to notice how low the levels of self esteem are in students, especially females in comparison to their parents and teachers.

The data was for various age groups in the same year. People who were 75+ in 1994 would have been born in the 1920's. Perhaps there are societal reasons at play.

Section 2d – Simulation and the Normal Distribution

Exploration: I will simulate a survey in which 15-19 year old females will respond to the question, "Do you have high self-esteem?" I will look at how the proportion of high self-esteem responses can vary between repeated random survey samples of one hundred 15-19 year old females.

Plan: As an estimate of the probability of getting a high self-esteem response, I used the same E-Stat data (table 104-0019) used in the probability section of the project.

Analysis: There were 369371 positive (high self-esteem) responses out of a total of 989002 females 15-19 year old. Thus the probability that a randomly surveyed 15-19 year old female will report high self-esteem is 0.373. Since this experiment can be defined in terms of success (equals high self-esteem response) and failure (equals not high self-esteem response), and if I assume that the individual responses are independent, then the experiment suggests a binomial probability model. With a sample of 100 females, the mean number of high self-esteem responses and the standard deviation is given by:

$$u = n*p \qquad\qquad \sigma = \sqrt{(npq)}$$
$$= 100*0.373 \qquad\qquad = \sqrt{(100*0.373*0.627)}$$
$$= 37.3 \qquad\qquad\qquad = 4.84$$

To construct this simulation I followed the instructions given in the Fathom Workshop Guide.

I put a slider on the fathom page set the slider to .373 and then used the random function to create a table of sample data.

probHSE = 0.373
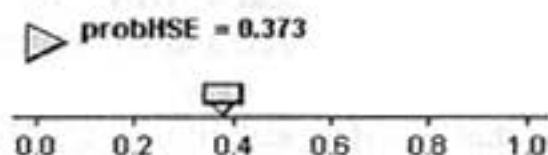
	Sample of 15-19 yr old F	
	Respon...	<new>
1	notHSE	
2	notHSE	
3	notHSE	
4	HSE	
5	notHSE	
6	HSE	
7	HSE	

Next I made a bar graph of the response distribution for the 100 females surveyed.
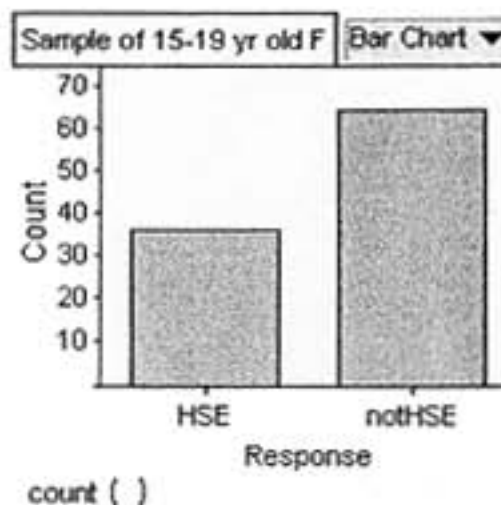
Then I repeated this simulation 100 times. The proportion of high self-esteem responses for each of the 100 simulations was recorded in a new, measures collection and charted as a histogram.

Measures from Sample of 15-19 yr old F

In a summary table I calculated the mean and standard deviation of the proportions gathered in the measures collection.

Measures from Sample of 15-19 yr	Histogram ▼
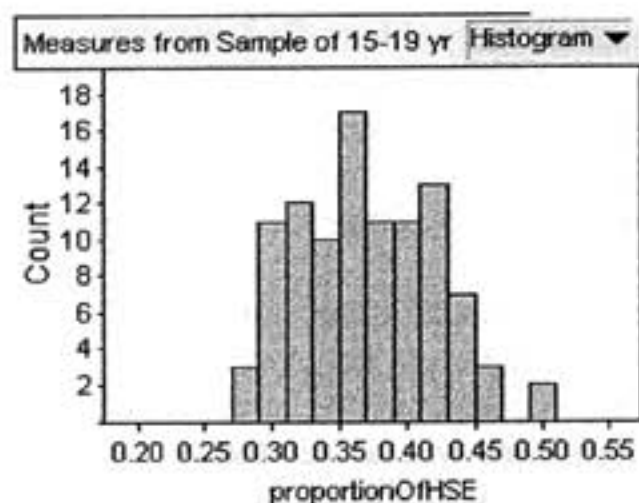
Measures from Sample of 15-19 yr old F	Summary Table
⇩ ⇨	**proportionOfHSE**
	0.3755
	0.050239326
	0.51
	0.28

S1 = mean ( )
S2 = stdDev ( )
S3 = max ( )
S4 = min ( )

Observations
In this simulation I found that the proportion of high self-esteem responses could vary a lot from sample to sample to sample (the highest proportion was 0.51 and the lowest was 0.28). However the mean proportion of all the samples, 0.3755 was close to the theoretical mean, proportion 37.3/100=0.373 for the binomial distribution.

Looking at the histogram, the distribution of the data accumulated in the simulation is approximately "bell-shaped and appears to follow a normal distribution. This is not surprising since a normal distribution is a reasonable approximation of a binomial distribution when $n \times p$ and $n \times q$ are both greater than 5 ($p = .373$ and $n = 100$ trials in my simulation, $q = 1-p$).

A normal probability or normal quantile plot gives a visual way to determine if a distribution is approximately normal. It is essentially a scatterplot of the actual data values plotted against the 'ideal' values from a normal distribution. The nearer that the points of the scatterplot lie to a straight line the more the distribution resembles a normal distribution. The normal quantile plot of my data shows a nearly straight line. Therefore my simulation data can be modelled by the normal probability distribution.

Measures from Sample of	Normal Quantile Plot ▼

Normal Quantile = 19.9proportionOfHSE - 7.5
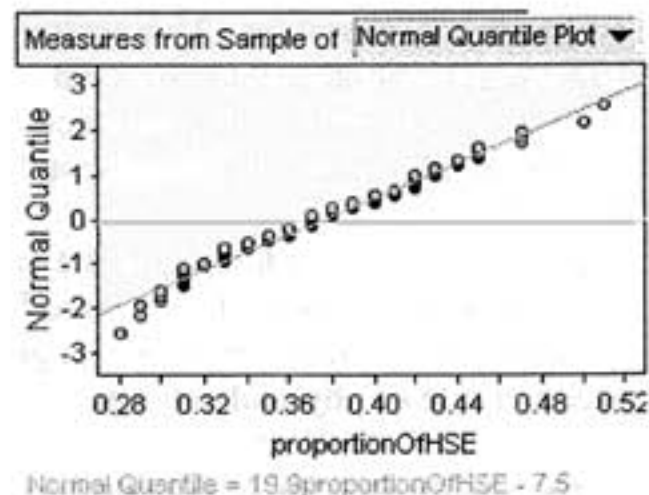
Project Conclusions

I began this project wanting to investigate if there was a relationship between participation in high school sports and a student's mental and/or physical health. While searching for data, mostly on the E-Stat site, I was amazed and sometimes frustrated by the mountains of data. The project ended up following directions based on the data available. Although I sometimes could not find exactly what I was looking for, I could usually find something that related in some way to my hypothesis. It did however; require a lot of time sifting though pages of data.

My particular data was restricted by the time frame the surveys were done in. Survey questions related to sports and self-esteem were only conducted in 1994, 1996 and 1998. It was not until the project section on probability that I was actually able to find raw data on my specific subject.

My research evolved through various stages until I finally ended up looking into a relationship between exercising 4 times a week and high self-esteem in males and females between the ages of 12-75+. The main focus, or age group of interest remained high school age Canadians.

I began by thinking that students who exercised 4+ times/week would likely have high self-esteem. I thought this would be true because; they would be busy, they would be healthy, and they would experience the benefits of exercise (i.e. confidence, stress relief, health). However, I was surprised to find that exercise frequency was not related to reporting high self-esteem. In fact, both males and females between the ages of 15-19 have a very low probability of reporting high self-esteem and a high probability of reporting exercising 4 times or more a week.

A lack of confidence is a common thing in teens; in fact it is the stage of life when high self-esteem is least likely. I wondered if teens that exercised frequently did so because of poor body concept, and that those who had high self-esteem didn't exercise because they were confident about the way they looked. (i.e. didn't feel too fat or too skinny) Teens as opposed to other age groups, probably report higher exercise 4+/week because of the natural activity level of young people. (school sports teams etc.) It was illustrated in the last graph in the one variable section of the project, how much of an effect sports participation had on the self-esteem of 12-15 year olds. In this case kids involved with sports teams seemed to show a much higher level of self- esteem, as well as having fewer problems with friends, and were able to see their friends more outside of school. This helped prove my original hypothesis but is not very strong because of the huge possibility of extraneous variables. This section also pointed out how much more active teen males are compared to females. The two variable section illustrated that there was a moderate positive linear relationship between 15 year old females who exercise 4-6 times a week and responded yes to the statement "I feel very happy about life" and "I never feel lonely".

Out of all my research, I found the data in the probability section of the project the most interesting, particularly the relative frequency graphs showing the probability of Canadians from nine different age groups being physically active. Females maintained a steadier rate of physical activity through their lives than the males whose levels dropped in mid life. I wondered why this would be. Perhaps men at this age are focusing on their careers more or have young families and don't have time to exercise. Their level of

activity picks up after the age of retirement, when they would have more free time. I wondered if women reported being more active because they are constantly under pressure from the media of having the ideal body. The relative frequency graphs of high self-esteem were also very interesting. I wonder why women peak in high self-esteem later than men. I also found it disturbing (yet obvious) that female's self-esteem drops in high school where males at this age quite high. There's an overall pattern of teens self-esteem being low while the age group of who would be their parents is quite high. This might explain some existing tension that most of us have felt at some point in our lives within the family.

In the simulation section I found that the frequency of high self-esteem responses could vary quite a bit from sample to sample. This simulation was very interesting but had some many outside factors that could sway the results that it was almost unrealistic. However it was successful in achieving a mean proportion of all the samples close to the theoretical mean.

So, does weekly exercise in high school students influence how they feel about themselves? Well I certainly found out from this project that exercise is not the only factor affecting self-esteem, however I was able to find proof and relationships in the some times negative and positive affect exercise has. Exercised proved to be beneficial in helping a child or teen meet friends or even feel important. Working out could also be indicator of wanting to change your body image, which could be unhealthy for teens. I find, in the end that it is difficult to conclude a definite relationship between exercise and the effect it has on self-esteem in high school students. Teenagers come from a wide variety of backgrounds, have different beliefs, and different opinions about exercise it is too difficult sum these up in a single answer.

Bibliography

http://estat_statcan.ca

Tables: 104-0019
         104-0033
         110-0005
         110-0018

National Longitudinal Survey of Children and Youth: Participation in Activities, 1989/99.

A Family Affair: Children's participation in sports

http://canada.gc.ca/publications./publication_e.html
Sport Participation in Canada, 1998 Report