

Using the  Statistics Canada / Statistique Canada

## Canadian Community Health Survey Microdata in MDM4U



OAME Annual Conference  
May 2009

Tricia Misner, Simcoe Composite School, GEDSB  
[patricia.misner@granderie.ca](mailto:patricia.misner@granderie.ca)

Shelah Pickett, North Park Collegiate, GEDSB  
[shelah.pickett@granderie.ca](mailto:shelah.pickett@granderie.ca)

Betty Temmer, Retired Math Head, GEDSB  
[bettytemmer@kwic.com](mailto:bettytemmer@kwic.com)

Yves Saint-Pierre Statistics Canada, Senior Education Outreach  
Consultant [mdm4u@statcan.gc.ca](mailto:mdm4u@statcan.gc.ca)

# Using the Canadian Community Health Survey Microdata in MDM4U

## What is CCHS?

CCHS stands for Canadian Community Health Survey. The survey, which is done every two years by Statistics Canada, contains over 2000 health and demographic attributes collected from a representative sample of approximately 130 000 Canadians age 12 and older. The CCHS microdata provides extensive information on the health of Canadians and on the Canadian health care system at the level of health region and province.

## Why do we need it?

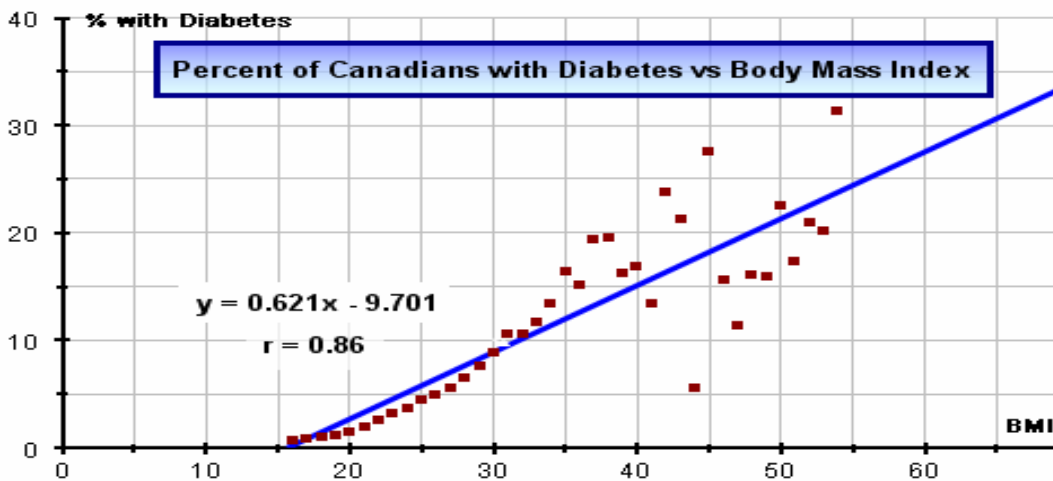
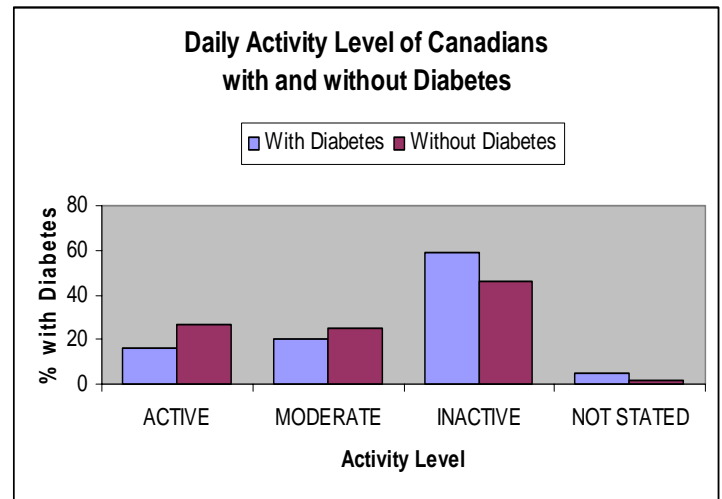
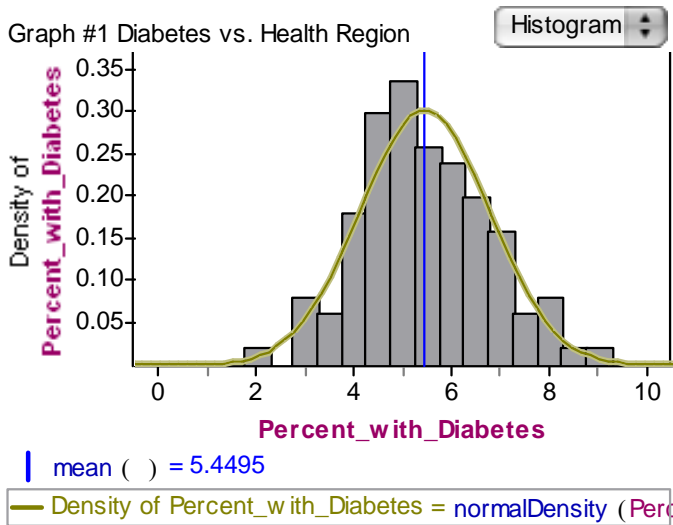
For students to be successful with the culminating investigation in the MDM4U course, they need a complex set of research, analytical, and organizational skills. Even after we feel we have given them these necessary skills, many students still do not know where to start or how to proceed. *Finding suitable data is one of the major stumbling blocks students encounter.* We have tried a variety of approaches to prepare the students for this task, but we feel that our current approach using the CCHS data from Statistics Canada is by far the most successful.

## Beyond 20/20 Browser

To extract the data, you will need to become familiar with the Beyond 20/20 software which is included with the data. Our goal today is to introduce you to this rich source of data and give you enough basic skills using the 20/20 Browser so that you will feel comfortable exploring further on your own. To do this, we will investigate the prevalence of diabetes in the health regions of Canada and several life style factors relating to diabetes. We will extract data on relevant variables, then graph and analyze the data using statistical analysis software such as Fathom, Tinkerplots, QuattroPro, Excel, or Autograph.

**We encourage you to invest a few hours to become familiar with the CCHS data files and the Beyond 20/20 Browser. The results are well worth it for both you and your students!**

# Sample of graphs made using CCHS



## Resources and Website Supports :

- [www.oame.on.ca](http://www.oame.on.ca) See link to narrated tutorials on how to use CCHS microdata with Beyond 20/20 Browser
- Statistics Canada has a new website address [www.statcan.gc.ca](http://www.statcan.gc.ca)
- <http://teacherweb.com/ON/statistics/Math/> - See Health Microdata Supports Folder for CCHS –specific teaching materials and student exemplars
- [www.statcan.ca/english/kits/courses/smath2.htm](http://www.statcan.ca/english/kits/courses/smath2.htm) - StatCan data for MDM4U
- Statistics Canada Education Outreach Consultants:  
For free workshops for your professional development day on any Statistics Canada resource (Census at School, E-STAT, function modeling or [www.statcan.ca](http://www.statcan.ca)), contact Sandra McIntyre [sandra.mcintyre@statcan.ca](mailto:sandra.mcintyre@statcan.ca) or [yves.saint-pierre@statcan.gc.ca](mailto:yves.saint-pierre@statcan.gc.ca)  
(Yves replaces Joel Yan who recently retired.)

## To become familiar with CCHS and Beyond 20/20, we will examine:



- a. Working with 2 variables: Percentage of *people with Diabetes* in the *Health Regions* of Canada. We will import the data to Fathom and display the results in a histogram.
- b. Working with 3 variables: Percentage of *people with Diabetes* in the *Health Regions* by gender (*sex*):  
We will produce a split histogram and box and whisker plot comparing the distribution of the % of females and males with diabetes.  
We will examine the relationship between the % of females with diabetes in a Region and the % of males with diabetes in the same Region by making scatter plot and adding a line of best fit.
- d. Compare diabetes rates with BMI and produce a scatter plot.
- e. Compare *diabetes* rates with *physical activity level*.  
We will look at this in 2 ways:  
How active are people with diabetes?  
What is the rate of diabetes for active and inactive people?
- f. Compare diabetes rates for different age groups. (Because *age* is grouped data, will need to add a new attribute in Fathom, MIDPOINT.)

## Some Useful Information:

### WHEN WORKING WITH 20/20, REMEMBER:

- Have only 2 variables active when changing to Worksheet and Distribution
- Don't hide unwanted fields until you have done your distributions or your percentages will not add to 100% Change decimal places from the VIEW menu if required
- When pulling up your variables using 20/20, think carefully which variable goes on the left and which goes on top.

## CCHS CONTENT TYPES

### Common content

These common content topics were asked of all respondents in all health regions.

### Optional content

Regions could select modules related to their particular needs and priorities. These are not easily generalized across Canada.

### **Sub-sample content**

These were asked in all regions but only of a subset of the respondents.

### **20/20 VOCABULARY**

*Dimension:* variable (Fathom calls these *attributes*)

*Extract:* a blank worksheet, before you hit the green light

*Table:* a worksheet with data after you hit the green light

### **CCHS DIMENSION NAMES**

*CCCE\_101* The CCC refers to the module (chronic care), the E refers to the cycle (this data is cycle 3.1, 2005) and the 101 refers to the question number in the questionnaire.

### **DOCUMENTATION**

There are 4 main areas in the Documentation which are of interest to us:

#### **User's Guide**

Here you will find lots of interesting background information including a listing of all the dimensions (p. 9- 11), sampling methods (p. 13), weighting methods (p.26), etc.

#### **Questionnaire**

This contains the exact questions that were asked of the respondents.

#### **Data Dictionary**

For each question, this gives the sample size for each Health Region where the question was question.

#### **Indices**

There is both a **Topical Index** and an **Alphabetical Index**.

*The following pages are a brief description of the process we use to teach students the skills they will need to do their investigation. The skills are taught by doing a mini-investigation on the topic of DIABETES  
These are the student handouts.*

## MDM4U Sample Culminating Investigation Topics on The Health and Wellness of Canadians (using CCHS)



1. Investigate the use of health care professionals by Canadians.
2. Investigate the satisfaction of Canadians with the health care system
3. Investigate if Canadians are interested in improving their health.
4. Investigate if Canadians have an active, healthy lifestyle.
5. Investigate the smoking habits of Canadians.
6. Investigate the consumption of alcohol amongst Canadians.
7. Investigate the use of illegal drugs amongst Canadians.
8. Investigate the home care services available to Canadians.
9. Investigate if Canadians are concerned about maintaining a healthy body weight?
10. Investigate the use of medications (prescription and non-prescription) by Canadians.
11. Investigate the problem of heart disease amongst Canadians.
12. Investigate the problem of asthma amongst Canadians.
13. Investigate the problem of arthritis amongst Canadians.
14. Investigate the problem of epilepsy amongst Canadians.
15. Investigate the problem of eye disorders amongst Canadians.
16. Investigate the problem of high blood pressure amongst Canadians.
17. Investigate the problem of migraine headaches amongst Canadians.
18. Investigate the problem of allergies amongst Canadians.
19. Investigate the problem of sleeping disorders amongst Canadians.
20. Investigate the problem of drinking and driving amongst Canadians.
21. Investigate the problem of depression amongst Canadians.
22. Investigate the problem of anxiety disorders amongst Canadians.
23. Investigate the problem of distress amongst Canadians.
24. Are pregnant Canadian women doing what's necessary to ensure the health and safety of their babies? Investigate.
25. Investigate the safe sex practices of Canadians.
26. Are Canadians safe drivers? Investigate.
27. Investigate the problem of stress in the workplace amongst Canadians.
28. Investigate the problem of cancer amongst Canadians.
29. Are Canadian women pro-active regarding female cancers (breast self-examination, mammograms, Pap smears)?
30. Investigate if second hand smoke is a health issue in Canada.
31. Are Canadians taking precautions to avoid injury (seatbelts, helmets, etc.) or are they reckless?
32. Investigate the problem of adequate medical/health insurance amongst Canadians.
33. Investigate the issue of flu shots by Canadians.
34. Investigate oral hygiene practices of Canadians (dental care, visits to dentists, etc.)

**ETC. : Have students search the Index for other topics!**

## MDM 4U Culminating Investigation

In your culminating investigation, you will

- Investigate a topic dealing with the health and well being of Canadians.
- Collect relevant background information for your topic
- Identify the variables you will investigate
- Gather data related to the study of the issue
- Interpret, analyze, and summarize your data using tables and graphs.
- Draw conclusions based on your statistical analysis.
- Compile a clear, well-organized, and detailed report of the investigation.
- Present a summary of the investigation (approximately 10 min) to the class using Power Point or Corel Presentation.
- Critique the mathematical work of others in a constructive manner.

### Part 1: Collecting Background Information about your Topic

Some excellent sites (there are many more!)

- Statistics Canada has a wealth of information as well as data. [www.statcan.ca](http://www.statcan.ca)
  - >Select Publications (under *Find statistics* in the menu on the left side)
  - >Select Free Internet Publications
  - >Select Health. There are many excellent resources here.
    - Some you may want to check: *Statistical Report on the Health of Canadians*
    - You can also find the entire CCHS 3.1 questionnaire here under *Health Reports*
- Other great websites:
  - Health Canada, Healthy Ontario* (find these with a Google search);
  - World Health Organization (WHO)* <http://www.who.int/en/>
  - CIHI.ca* (Canadian Institute for Health Information)
  - EatRightOntario.com*
  - mirror-mirror.org* (on eating disorders)
  - [www.Canadian-Health\\_Network.com](http://www.Canadian-Health_Network.com)
  - Human Development Reports (This is a UN site, EXCELLENT info!)
  - The Daily* at [www.statcan.ca](http://www.statcan.ca) homepage

### Part 2: Finding Variables using The CCHS Database

Most of your data will come from the Canadian Community Health Survey (CCHS). To learn how to use this Data Base, we will work together through a mini sample project on *diabetes*. This sample project will be part of the process mark for your investigation.

To access the CCHS 3.1 data on any computer at BCI:

Go to: sdrive > MDM4U > CCHS2005\_Distribution >Index.htm >English >Accept

Finding relevant variables

- CCHS has an *Alphabetical Index*, a *Topical Index*, and a *Questionnaire* containing a full list of all questions asked for each variable.
- To find relevant variables for your topic: In the menu box on the left side, choose *Documentation* >Index. A page will open up in *Adobe*. Click on the binoculars in the tool bar. This gives you a SEARCH BOX.

## Part 3: Creating Graphs and Forming Conclusions from CCHS Data Using Beyond 20/20

*Beyond 20/20* is the programme used to extract data from the CCHS database.

As we learn to use it, it is important that you make any additional notes that you will need in order to remember how to work with this programme.

You can download this programme for free at

<http://estat.statcan.ca/Estat/English/beyond20.shtml>

BELOW ARE BRIEF INSTRUCTIONS FOR OBTAINING DATA USING 20/20 and CCHS. ADD TO THE INSTRUCTIONS AS NECESSARY AS WE WORK THROUGH THIS TOGETHER IN CLASS.

Skills you will need:

- Finding variables using *Data > all > search*
- Making a 20/20 Table with 2 or more variables
- Changing your data into percentages instead of frequency
- Importing your data into Excel or Fathom

- Open CCHS. In the box on the left, choose *Data*

- Near the bottom of this page choose

[Beyond 20/20 Browser](#) (select **Open**) – for common and optional content

- For variables, go to DATA in menu bar > find source fields > choose *all* and *search*

- When you have your variables, select the *green light* in menu bar

-Go to VIEW in menu bar to change to *worksheet* (select all variables)

- Go to VIEW again and change to *distribution* (select all variables, and ONLY row distribution, unselect all other choices).

### **IMPORTANT:**

- Only have 2 variables active when changing to *Worksheet* and *Distribution*
- Don't hide unwanted fields until you have done your *distributions* or your percentages will not add to 100%
- Change decimal places from VIEW menu if required

## SAMPLE Diabetes Investigation

Save all diabetes graphs as Fathom Files in your *Diabetes Folder* for Assignment # 2. Remember that all graphs must have a SOURCE reference, so keep track of the variables you are using.

### GRAPH 1, 2, and 3

Histograms and Box and Whisker Plots:

- a. Graph 1: Percentage of people with Diabetes in the Health Regions of Canada  
Import to Fathom. Clean up data if required. Rename attributes and collection.  
Produce histogram  
Add mean, and standard deviation. What do these tell us about the prevalence of diabetes in different areas of Canada?  
Locate Brant Health Region on your graph.  
Note that this data is approximately normally distributed.
- b. Graph 2: Working with 3 variables: percentage of people with Diabetes, Health Regions, and gender:  
Produce a split histogram showing the distribution of the % of females with diabetes and the distribution of the % of males with diabetes. Compare using statistical analysis.
- c. Graph 3: Produce a split box and whisker plot for the variables in part b.

**IMPORTANT:** - If you have more than 2 variables, when doing distributions, only 2 can be active, the others must be in the holding area.

### GRAPH 4 & 5 & 6

Scatter Plots:

- a. Graph 4: Using the same table as above, we want to determine if there a relationship between the % of females with diabetes in a Region and the % of males with diabetes in the same Region. Produce a scatter plot. Add a line of best fit. Determine  $r$ . What conclusion can you draw?
- b. *NOTE: When pulling up your variables using 20/20, think carefully which variable goes on the left and which goes on top.*

Graph 5: Compare diabetes rates with BMI. (You will need to do lots of "cleanup" with this data.) To find this variable, search *BMI* and choose the one that says *numerical*

Graph 6: Compare diabetes rates with age. For this one you will need to add a new attribute, MIDPOINT  
Add a regression line to both graphs and determine the correlation coefficient.  
What conclusions can you draw?

## GRAPH 7, 8 & 9

Bar graphs...we will make these using EXCEL (or QuattroPro)

- a. Graph 7: Physical activity level vs. Diabetes by sex; Search for *Activity* and use the variable *Activity Index*. Make a double bar graph showing the percentage of males and females with diabetes for each activity level.
- b. Graph 8: Diabetes vs. Physical Activity Level. Use the same variables as in graph 7 but this we want to know the level of physical activity of people with diabetes. So, *Diabetes* goes on the left, *Activity Index* goes on top.
- c. Graph 9: Vision vs. diabetes

### **NOW...** FINISHING ASSIGNMENT # 2

You should have 9 graphs in your diabetes Folder.

- clean them up and make them into **"WOW GRAPHS"**!
- Copy all graphs into a Word/Word Perfect document
- organize them (number them, titles etc.)
- Pick 2 graphs and write a brief introduction and conclusions based on your analysis.

AND NOW YOU ARE READY TO START GATHERING CCHS DATA AND CREATING GRAPHS FOR YOUR OWN INVESTIGATION USING CCHS AND OTHER SOURCES.....

## **Part 4: Finding Health Related Information and Statistics on the StatCan website**

- A. From the StatCan Homepage, at [www.statcan.gc.ca](http://www.statcan.gc.ca)**

### **Summary Tables**

Summary tables, provide continuously updated national and provincial data gathered through hundreds of surveys. These top 500 tables are sorted into 32 broad subject categories. You can search the tables by subject, province, territory or metropolitan area.

1. > Tables by subject >search summary tables (at the top) *diabetes*  
> choose No. 2, persons with diabetes by age and sex > choose % in upper left corner - you now have a time series for % of Canadians over 12 with diabetes
2. Go back to "search summary tables *diabetes*" and choose No.3, *Selected leading causes of Death*. At the top of the page under *Related Tables*, choose *Measures of Health*: Check out all the tables listed here. Choose *BMI changes*. At the bottom of the grey area, choose CANSIM 1047030. Download as a Beyond 20/20 table. Hide everything except *Percent*. Notice all the variables in the holding area at the top. Try dragging the BMI categories to the left....

REMEMBER: You can download the Beyond 20/20 Browser for free at <http://estat.statcan.ca/Estat/English/beyond20.shtml>

## The Daily

The Daily news releases cover a wide range of topics and usually include tables and/or graphs. Use The Daily's search engine to find articles of interest and then sort using date to find the most recent ones. Articles often refer to CANSIM table numbers that allow you to further investigate the data. Find these tables free in E- STAT. Articles also link to a section called 'Definitions, data sources and methods: survey number', where you can read about sampling techniques and view the actual questionnaire.

## Health Indicators Database

StatCan has collected many health related documents and data into a single database. This includes publications, articles, and statistics. It includes over 80 indicators which measure the health of the Canadian population and the effectiveness of the health care system

To access: from the home page, search for *Health Indicators*.

From the list, choose *1. Health Indicators: Product Main Page*.

Scroll down to Subjects - Health. EXPLORE these links!

Note also the list of "keywords" to help you in searching for specific variables.

## B. E-STAT

Developed specifically for schools, E- STAT contains chronological data from hundreds of social and economic surveys, as well as several hundred census variables for every province, county and municipality in Canada, right down to census tracts or neighbourhoods

(Remember: at home you need user name: *granderie*, password: *estat*)

CANSIM will give you Time Series for thousands of Variables. This data is FREE through E-STAT but some of it costs \$ if you access it through the main StatCan site.

CHECK OUT, from the choices on the left

**Table of Contents** >Health ... check out all the options

At bottom of page, choose *Health Behaviours of School-Aged Children*

Explore these tables. Pick one and download as a 20/20 file

**Search CANSIM.** Try searching for pregnancy, or suicide, or any other variable you might be interested in.

**Other Sources of Data:** Check out these sites

- **World Health Organization (WHO)** <http://www.who.int/en/>  
United Nations specialized agency for health.
- **Human Development Reports**  
This is a UN site - EXCELLENT
- [nationmaster.com](http://nationmaster.com)  
Stats on many countries
- [InfoNation](http://Infonation.com)  
A UN site that allows you to compare data from all UN members up to 5 countries at a time
- [www.2ontario.com](http://www.2ontario.com)  
Info on all aspects of Ontario
- [www.ec.gc.ca](http://www.ec.gc.ca)  
Environment Canada
- [www.tc.gc.ca](http://www.tc.gc.ca)  
Transport Canada
- <http://www.gdsourcing.com>  
Links to many useful sites

## **Part 5: The Final Product -COMPILING YOUR REPORT -**

Your Report should contain:

- A cover page
- An introduction to your topic. This should include relevant background information and a brief description of the variables you are examining. You should have approximately 8 variables related to your topic. Introduction should not exceed a page. To avoid plagiarism, you must acknowledge all information that you get from other sources. (Information on this is on the BCI Homepage under Library >APA Style Guide.)
- For each variable:
  - introduce the variables, provide a graph, statistical analysis and conclusions based on the analysis, not merely your opinion. You may include charts, diagrams, summary tables, etc. in your discussion of the variables.  
(One variable graphs: bar graphs, circle graphs, histograms; two variable graphs: scatter plots.) For one variable statistics, did you find measures of central tendency and measures of spread where applicable? For two variable analysis, did you use the line of best fit to classify the correlation and identify/predict a trend? What conclusions can you make based on these? Make sure these graphs are labeled clearly and have the source cited.
- At the end of the document, make a general conclusion on your topic based on your findings.
- A Bibliography or Works Cited page (Information on this is on the BCI Homepage under Library >APA Style Guide.)
- A rubric for assessing the report will be provided.

## **Part 6: The PRESENTATION**

- Prepare a Power Point of the key ideas in your report to present to the class
- Length: approximately 10 minutes.
  - The 10 minutes includes time for questions and answers on your topic.
- A rubric for assessing the presentation will be provided

## **Part 7: Critiquing Your Peers**

- Critique the work of your peers in a constructive manner.

## MDM 4U Culminating Investigation: EVALUATION

### Knowledge/Application

Criteria	Level R	Level 1	Level 2	Level 3	Level 4
Introduction gives relevant background information and Introduces variables to be examined	with no effectiveness	with limited effectiveness	with some effectiveness	with considerable effectiveness	with a high degree of effectiveness
Graph Creation - suitable scale - well labeled - Source given	No acceptable graphs created	Few acceptable graphs constructed; source/axes not clear	Some acceptable graphs; some clarity in labeling	Most graphs are reasonable and labels are clear	All graphs are constructed properly and are clearly labeled
Variety of graph types included (histogram, scatter, bar, etc.) Correct choice of graph type	No variety used	Limited variety used with	Some variety used type sometimes chosen appropriately	Considerable variety used, type usually chosen appropriately	A wide variety of graph types used with a high degree of effectiveness
Mathematical procedures are correct, appropriate, and complete	rarely	occasionally	sometimes	usually	always

### Critical & Creative Thinking/Communication

Criteria	Level R	Level 1	Level 2	Level 3	Level 4
Choice of Variables	chosen with no effectiveness	chosen with limited effectiveness	chosen with some effectiveness	chosen with considerable effectiveness	chosen with a high degree of effectiveness
Graph Analysis using appropriate 1 and 2 variable statistical methods	No analysis provided Conclusions from data are not valid	A minimal amount of analysis, errors in conclusions	Some graphs are analyzed; some validity in conclusions	Most graphs are analyzed and most conclusions are valid and supported; some depth	All graphs are analyzed and conclusions are justified, valid and insightful
Terminology and notation are correct	rarely	occasionally	sometimes	usually	always
Report Compilation	No evidence of logical organization and correct use of mathematical and grammatical conventions	Minimal evidence of logical organization and correct use of mathematical and grammatical conventions	Some evidence of logical organization and correct use of mathematical and grammatical conventions	Uses logical organization and correct use of mathematical and grammatical conventions	Consistently uses logical organization and correct mathematical and grammatical conventions
Conclusion made based on findings	with no effectiveness	with limited effectiveness	with some effectiveness	with considerable effectiveness	with a high degree of effectiveness
Bibliography/ works cited	rarely	occasionally	sometimes	usually	always

THERE WILL BE 2 SEPARATE AREAS OF EVALUATION FOR YOUR INVESTIGATION: PROCESS and PRODUCT

- The **process component** will be part of your term mark. It will consist of the assignments below.
- The Final **Report and the Presentation** will make up half of the 30% summative mark for the course. It is therefore worth 15 % of the course mark. A detailed rubric is provided.

## PROCESS ASSIGNMENTS

**Assignment 1.** DUE \_\_\_\_\_

- Background information collected.
- Rough draft of introduction written. This may be in point form at this time
- Mind Map using *Smart Ideas* showing your areas of investigation and the variables you are investigating. Submit the mind map as an appendix with your project.

**Assignment 2.** DUE \_\_\_\_\_

Open a folder entitled Diabetes Investigation. Save all graphs produced in class on *Diabetes* in this folder.

This will be submitted and given a completion mark.

**Assignment 3.** DUE \_\_\_\_\_

Submit 2 graphs (different types) that you have produced that you think show relevant information about your topic. Include a statistical analysis using methods we learned.

(For 1-Variable data: measures of central tendency and spread, standard deviation; for 2-variable data: trends/patterns using line of best fit, correlation coefficient, etc.)

**Assignment 4.** Due \_\_\_\_\_

Produce at least 2 Power Point (or Corel Presentation) slides showing your opening/introduction.

## CULMINATING INVESTIGATION DUE DATES

- Presentations will start on \_\_\_\_\_. Order of presentation will be by lottery.
  - Your Power Point PRESENTATIONS is due AT LEAST **TWO DAYS PRIOR TO your presentation date.**
  - 
  - All written REPORTS are due on the last day of presentations, \_\_\_\_\_.
  - You may email me your PRESENTATION or submit it on a CD
- IMPORTANT: You are expected to do most of the work in your own time.

CLASS WORK PERIODS: \_\_\_\_\_