Course Description:

Students will apply methods for organizing large amounts of information; apply counting techniques, probability, and statistics in modelling and solving problems; and carry out a culminating project that integrates the expectations of the course and encourages perseverance and independence.
Mark Breakdown:

Course Work (70%)

Tests (50%)
Assignments (20%)

Summative (30%)

ISU (10%)
Final Exam (20%)

Overview of Ongoing ISU Project

1. You will pose a significant problem whose solution would require the organization and analysis of a large amount of data.

2. You will apply the skills you learn in the course to design and carry out a study of the problem.

3. Compile a clear, well-organized, and fully justified report of the investigation and its findings.

4. Present your findings to the class in a seminar.
1.1 Visual Displays of Data

**Classroom Expectations**

RESPECT

SUCCESS

Washroom

Cell phones and IPods

List as many careers as you can think of that use data management:
1.1 Visual Displays of Data

Here are a few examples:

- investment advisor
- clinical research associate
- lawyers
- psychologists
- media/advertising
- statisticians
- meteorologists
- tourism
- insurance
- credit card companies

Watch how data management techniques can be used:

http://www.youtube.com/watch?v=jbkSRLYSsjo
Section 1.1 - Constructing and Interpreting Visual Displays of Data
"The Pittsburgh Penguins are the most popular team in the NHL"

In order to draw conclusions, like the one above, information must be gathered, organized, and displayed clearly.

**Tables** are used to organize data  
- Frequency tables and stem and leaf plots

**Graphs** are used to display data  
- Histograms, bar graphs, pictographs, circle graphs, box and whisker plot, broken-line graph, scatter plot

Definitions you need to know…..

**Population** - refers to the entire group about which data are being collected.

**Data** - information providing the basis of a discussion from which conclusions can be drawn (often takes the form of numbers)

**Sample** - part of a population that is selected to gain information about the whole population

**Frequency** - the number of times an event occurs or the number of items in a given category

**Frequency Table** - a table listing a variable together with the frequency of each value
Census vs. Sample

When information is gathered from all people in a population, the activity is called a ___________ of the population.

A poll (or survey) is a method of collecting data from a ___________ of the population.

Why use a sample?

- entire population is not always available
- time

Part 1: Organizing Data Using Tables
Example 1 - Frequency Table

Grade 12's were asked when their spares were and these were the results:

A, B, C, D, A, D, D, B, A, C, A, C,
B, B, B, A, D, C, A, A, B, D, C, A,
B, A, C, C, D, A, B, A, B, B

a) Organize it into a frequency table
b) How many students responded?
c) What percent of Grade 12's have their spare during block B?

<table>
<thead>
<tr>
<th>Spare</th>
<th>Tally</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>######1</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>######</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>####11</td>
<td>7</td>
</tr>
<tr>
<td>D</td>
<td>####1</td>
<td>6</td>
</tr>
</tbody>
</table>

b) 34

c) \[
\% \text{B spare} = \frac{\# \text{B spare}}{\# \text{surveyed}} \times 100 \\
= \frac{10}{34} \times 100 \\
= 29.4\% 
\]
1.1 Visual Displays of Data

Example 2 - Frequency Table

a) Class survey of favourite Atlantic Division NHL team?

<table>
<thead>
<tr>
<th>Team</th>
<th>Tally</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penguins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rangers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Devils</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islanders</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flyers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b) What percentage of students chose Pittsburgh?
Example 3 - Stem and Leaf Plot and Frequency Table With Intervals

The heights of grade 9 students were measured and recorded below:

- 131, 145, 151, 175, 181, 185, 187,
- 161, 130, 125, 129, 112, 110, 188,
- 185, 133, 122, 188, 151, 175, 167,
- 164, 174, 181, 156, 138, 189, 182,
- 111, 171, 121, 122, 175, 143, 139,
- 156, 181, 121, 137, 189, 144, 133

a) Display this data using a stem-and-leaf plot.
b) Display a Frequency table using INTERVALS
c) Why are intervals useful?

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>012</td>
</tr>
<tr>
<td>12</td>
<td>112259</td>
</tr>
<tr>
<td>13</td>
<td>0133789</td>
</tr>
<tr>
<td>14</td>
<td>345</td>
</tr>
<tr>
<td>15</td>
<td>1166</td>
</tr>
<tr>
<td>16</td>
<td>147</td>
</tr>
<tr>
<td>17</td>
<td>14555</td>
</tr>
<tr>
<td>18</td>
<td>11125578899</td>
</tr>
</tbody>
</table>
A reasonable class interval for this data is a spread of 10 units. Given that the smallest value is 110 and the largest value is 189, the intervals to best display this data should are 110-119, 120-129, and so on.

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>110-119</td>
<td>3</td>
</tr>
<tr>
<td>120-129</td>
<td>6</td>
</tr>
<tr>
<td>130-139</td>
<td>7</td>
</tr>
<tr>
<td>140-149</td>
<td>3</td>
</tr>
<tr>
<td>150-159</td>
<td>4</td>
</tr>
<tr>
<td>160-169</td>
<td>3</td>
</tr>
<tr>
<td>170-179</td>
<td>5</td>
</tr>
<tr>
<td>180-189</td>
<td>11</td>
</tr>
</tbody>
</table>

Class Interval: a category or division used for grouping a set of observations

Using individual items would create a table with the data so spread out that it would become difficult to view any trends
Example 4 - Stem and Leaf Plot and Frequency Table With Intervals

The points for the 30 NHL teams from the 2013 regular season are recorded below:

72, 56, 55, 49, 48, 63, 62, 57
56, 48, 57, 51, 42, 40, 36, 77
60, 56, 55, 41, 59, 55, 45, 42
39, 66, 59, 57, 51, 48

a) Display this data using a stem-and-leaf plot.
b) Display a Frequency table using INTERVALS
c) What percentage of teams had 70 points or more?

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>69</td>
</tr>
<tr>
<td>4</td>
<td>012258889</td>
</tr>
<tr>
<td>5</td>
<td>1155566677799</td>
</tr>
<tr>
<td>6</td>
<td>0236</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
</tr>
</tbody>
</table>

The tens digits are called the stems
The unit digits are called the leaves
1.1 Visual Displays of Data

b)

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 - 39</td>
<td>2</td>
</tr>
<tr>
<td>40 - 49</td>
<td>9</td>
</tr>
<tr>
<td>50 - 59</td>
<td>13</td>
</tr>
<tr>
<td>60 - 69</td>
<td>4</td>
</tr>
<tr>
<td>70 - 79</td>
<td>2</td>
</tr>
</tbody>
</table>

c) \[
\% \geq 70 = \frac{\# \geq 70}{\# \text{ teams}} \times 100
\]

\[
= \frac{2}{30} \times 100
\]

\[
= 6.7\%
\]

Part 2: Displaying Data Using Graphs
Bar Graph Vs. Histogram

Bar graphs are generally used to represent nominal/discrete data.

For example: days of the week, months of the year, types of chocolate bars, NHL teams, number of heads when you flip a coin, number of threes when you roll a die, number of people sleeping in a stats class.

A bar graph consists of parallel bars of equal widths (with a space between each bar) with lengths proportional to the frequency of the variables they represent.

Discrete Variable: a variable that can take on only certain values within a given range. (whole numbers)

Example 1 - bar graph

a) Conduct a class survey of birth months and record the results in a frequency table

b) Should you use a bar graph or histogram to graph the data? Why?

c) Graph the data

d) What percentage of students in this class were born this month?
1.1 Visual Displays of Data

a) 

<table>
<thead>
<tr>
<th>Birth Month</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>3</td>
</tr>
<tr>
<td>February</td>
<td>2</td>
</tr>
<tr>
<td>March</td>
<td>4</td>
</tr>
<tr>
<td>April</td>
<td>6</td>
</tr>
<tr>
<td>May</td>
<td>2</td>
</tr>
<tr>
<td>June</td>
<td>0</td>
</tr>
<tr>
<td>July</td>
<td>2</td>
</tr>
<tr>
<td>August</td>
<td>1</td>
</tr>
<tr>
<td>September</td>
<td>2</td>
</tr>
<tr>
<td>October</td>
<td>3</td>
</tr>
<tr>
<td>November</td>
<td>0</td>
</tr>
<tr>
<td>December</td>
<td>1</td>
</tr>
</tbody>
</table>

b) 

bar graph;

months of the year is nominal data
c) Birth Months of Stats Class

![Bar chart showing birth months]

\[
\text{% born Sep} = \frac{\# \text{ born Sep.}}{\# \text{ Students}} \times 100
\]

\[
= \frac{2}{26} \times 100
\]

\[
= 7.7\%
\]
1.1 Visual Displays of Data

Bar Graph Vs. Histogram

A histogram is a frequency distribution where the horizontal access is divided into equal class intervals in to which data have been divided. The heights of the rectangles (that have no spaces between them) represent the frequencies associated with the corresponding intervals. It is important that each interval have the same width. Histograms are typically used for continuous data.

For example: height of students, speed of a car, weight of a person, time to wake up in the morning.

Continuous Variable: a variable or data that can have an infinite number of possible values in a given interval. A measure of quantity will always be continuous.

Frequency Distribution: a set of values of a variable, together with the frequency of each value.

Example 2

a) Is the following graph a bar graph or a histogram? How do you know?

Histogram because there is no space between the bars.
b) Which height interval has the highest frequency? What is the frequency?

- 180-190 cm
- The frequency is 17.

**How to Make a Histogram**

1. Choose the number of intervals (if the question doesn’t specify, choose between 5 and 10)

2. Calculate the range of your data (largest data point – smallest data point)

3. Round your range UP to a number that is easily divided by the number of intervals you chose.

4. Calculate your bin width \( \text{bin width} = \frac{\text{range}}{\text{number of intervals}} \)

5. Determine the first value for your first interval. \( \text{lowest value} - \frac{\text{rounded range-actual range}}{2} \)

6. If any data points fall on the border of any of the intervals, add a decimal place to ensure that this doesn’t happen.

7. Make a frequency table using the intervals you have determined.

8. Draw the histogram (no spaces between bars)
1.1 Visual Displays of Data

Example 3 - histogram

Here are a class' scores obtained on a data management exam:

78, 81, 55, 60, 65, 86, 44, 90
77, 71, 62, 41, 80, 72, 70, 64
88, 73, 61, 70, 75, 98, 51, 73
59, 68, 65, 81, 78, 67

a) Determine the range of the data

\[
\text{range} = \text{largest data point} - \text{lowest data point}
\]

\[
= 98 - 41
\]

\[
= 57
\]

b) Determine an appropriate bin (interval) width that will divide the data into 6 intervals.

\[
\text{bin width} = \frac{\text{range}}{\text{number of intervals}}
\]

\[
= \frac{57}{6}
\]

\[
= 10
\]

Note:
Round your range UP to a value that can be divided easily.
1.1 Visual Displays of Data

c) Determine the first value of your first interval

We added $\frac{3}{2}$ to 57 when we rounded our range, therefore we should subtract $\frac{3}{2} = 1.5$ from our smallest value $41$; which makes our starting point $39.5$.

or just use the formula:

$$\text{initial value} = 41 - \left( \frac{60 - 57}{2} \right)$$

$$= 41 - \frac{3}{2}$$

$$= 39.5$$

Note:

1. If you have rounded your range up you should subtract half of the amount you rounded from the smallest value to evenly distribute the 'excess of your range'.

2. Make sure no data points lie on the border of two intervals. (Do this by subtracting .5 from a whole number, .05 from data with one decimal point, .005 from data with two decimal points and so on)

d) Create a frequency table using your intervals

<table>
<thead>
<tr>
<th>Grade Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>39.5 - 49.5</td>
<td>2</td>
</tr>
<tr>
<td>49.5 - 59.5</td>
<td>3</td>
</tr>
<tr>
<td>59.5 - 69.5</td>
<td>8</td>
</tr>
<tr>
<td>69.5 - 79.5</td>
<td>10</td>
</tr>
<tr>
<td>79.5 - 89.5</td>
<td>5</td>
</tr>
<tr>
<td>89.5 - 99.5</td>
<td>2</td>
</tr>
</tbody>
</table>

Notice that the number one interval ends with, the next interval starts with the same number. This is because the data for a histogram is continuous!!!
Example 4 - Pictograph

A pictograph is a symbolic representation of data. The following pictograph displays the number of participants, aged 15 and older, in the five most popular sports activities in Canada.

<table>
<thead>
<tr>
<th>Sport Activity</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golf</td>
<td>🤛前十人</td>
</tr>
<tr>
<td>Ice Hockey</td>
<td>🤛前十人</td>
</tr>
<tr>
<td>Baseball</td>
<td>🤛前十人</td>
</tr>
<tr>
<td>Swimming</td>
<td>🤛前十人</td>
</tr>
<tr>
<td>Basketball</td>
<td>🤛前十人</td>
</tr>
</tbody>
</table>

Legend: 🤛 represents 100,000 people

How many people (aged 15 and older) play hockey? 

1 500,000
1.1 Visual Displays of Data

Example 5

<table>
<thead>
<tr>
<th>Varieties of Apples in a food store</th>
<th>How many Red Delicious apples are in the store?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Delicious</td>
<td>30</td>
</tr>
<tr>
<td>Golden Delicious</td>
<td></td>
</tr>
<tr>
<td>Red Rome</td>
<td></td>
</tr>
<tr>
<td>McIntosh</td>
<td></td>
</tr>
<tr>
<td>Jonathan</td>
<td></td>
</tr>
</tbody>
</table>

How would you represent 11 apples?

Problems with Pictographs: they can make a graph more interesting but relative frequencies for the categories can be hard to read sometimes.

Circle Graphs

Here is a circle graph for goals scored by Pittsburgh Penguins from the 2013 season:
A circle graph or pie chart is a circle divided into sectors whose areas are proportional to the quantities represented.

<table>
<thead>
<tr>
<th>Mode of Transportation</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike</td>
<td>10</td>
</tr>
<tr>
<td>Walk</td>
<td>15</td>
</tr>
<tr>
<td>Bus</td>
<td>9</td>
</tr>
<tr>
<td>Car</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
</tr>
</tbody>
</table>

Box and Whisker Plot

In a box and whisker plot, the box contains the median of the data and its width represents the middle half of the data.

The upper and lower limits for the box are found by finding the median for the upper and lower half of the data.

From the sides of the box, horizontal lines are drawn extending to the highest and lowest data points.
Example 7

A random survey of people at a golf course asked them how many times they had seen Happy Gilmore. The results are shown below in ascending order

1, 2, 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 8, 9, 10, 10, 12, 15, 26
1.1 Visual Displays of Data

a) Find the median of the data ($Q_2$).

The median is either the middle of the set of data, or, the mean of the two middle data (if there is an even number of data points).

b) Find the median of the lower half of the data ($Q_1$).

c) Find the median of the upper half of the data ($Q_3$).

d) Create a box and whisker plot for the data.
1.1 Visual Displays of Data

**Broken Line Graph**

Broken line graphs are useful for noticing trends in continuous data. They are created by joining data points with line segments.