CPSC 203 Spreadsheets

Week 2 Lab2

Spreadsheet Design Rules

Dina A. Said
dasaid@ucalgary.ca

Based on:
Lecture Notes of Dr. Jalal Kawish

Goals for Today

- Practice visual elements
- Get Familiar with basic design rules

- Next Time:
  - Revision for the first Quiz.

Basic skills you will need

- To reference a cell D2 in another sheet e.g. Data
  - Data!$D$2
- To reference a range D2:D9 in another sheet e.g. Data
  - Data!$D$2:$D$9
- Some keyboard shortcuts:
  - Copy: ctrl + c
  - Cut: ctrl + x
  - Paste: ctrl +v
  - Select all: ctrl+ a
  - Undo: ctrl + z
  - Redo: ctrl + y

Basic Design Rules

- Design the spreadsheet on paper first.
- Test and edit your calculations.
- Keep the components of a calculation visible.
  - For example: The reader should not go over the cell to know the number you multiplied by.

Basic Design Rules (cont.)

- Be aware of the "space" or "geography" of the spreadsheet. Arrange your information so that it is well spaced and easy to take in at a glance.
  - You can use as many sheets as you can
  - You should give meaningful names to columns, rows, and sheets.
  - Consider *wrap text, shrink to fit, and merge cells* for better visualization for texts

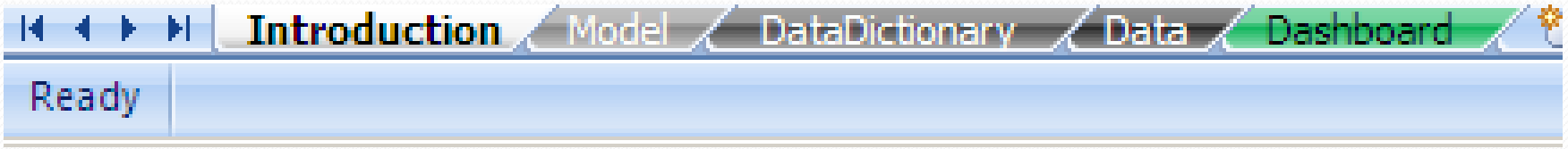
Spreadsheet Properties

- Good character
- Easy to build
- Easy to read
- Easy to use
- Easy to change
- Error free

Contents of a Spreadsheet

1. Introduction:
   - Introduction, title, description, and contents
2. Model:
   - Main form data
3. Data Dictionary
   - Explains columns and calculations
4. Data:
   - Data used in your sheet
5. Dashboard
   - Visual reports (charts)

Tabs



Contents of a Spreadsheet

- Introduction:
  - Introduction, title, description, and contents
- Model:
  - Main form data
- Data Dictionary
  - Explains columns and calculations
- Data:
  - Data used in your sheet
- Dashboard
  - Visual reports (charts)

Intro – 1. Make a formal intro

- Intro section should give the user a clear idea of how the sheet ties with the real world
- Intro devices:
  - Include a title that passes critical info
  - Declare the purpose
  - Give directions on how to use the model
  - Include references
  - Include a table of contents

Who Are We?

CPSC 203 L60 Assignment 1

Any group of people, regardless of size, can be noted for its diversity. Every single person has their own identity, their own skills, strengths,weaknesses, interests, and dislikes. This computer science class is no exception.

Category I - Most Visited Websites

Each person in the class was asked which 5 websites they visited the most. The results were then tabluated and can be found in the 'Data' tab. The most popular websites were placed on a chart showing just how popular they were, which can be found in the 'Dashboard' chart (Figure 1).

Category II - Internet Interests

Each person in the class was asked to rate how often they perfomed a number of activities on the Internet on a scale from 0 to 5, where 0 was 'do not use at all', and 5 was 'use extremely often'. The activities included shopping, gaming, Email, Research and Searching, Music Downloads, friends, and photography.

Category III - Computer Experience

Each person in the class answered whether or not they had used various computer technology products in their lives, including Computers, Windows PC's, Macs, IRC, LinuxBox, Excel, Word, OpenOffice, Email, Skype, Wikipedia, Access, and Powerpoint.

Would you like instructions on working with this Excel file?

>>>To view the raw data, and to see calculations done to it, click the 'Data' tab.
>>>To view how the ideas for the charts and the formulas were put together, click the 'Model' tab.
>>>To better understand any calculations made in this file, click the 'DataDictionary' tab.
>>>To view the charts in an Information Dashboard setting, click the 'Dashboard' tab.\

This and the following examples are from an **older** 203 assignment

Intro – 2. Informative Title

- Title must at least include:
  - Name of the model
  - Date
  - Name of creator
- Properties of a title:
  - Short
  - Apt (to the point)
  - Memorable

Intro – 3. Declare the Purpose

- Tell the spreadsheet user what the spread sheet is all about

- For instance, the purpose of your assignment 1 spreadsheet could be:

Analyzing movie preferences and watching habits of the computer science 203 class, namely blah blah …

Example Intro

Category I - Most Visited Websites

Each person in the class was asked which 5 websites they visited the most. The results were then tabluated and can be found in the 'Data' tab. The most popular websites were placed on a chart showing just how popular they were, which can be found in the 'Dashboard' chart (Figure 1).

Category II - Internet Interests

Each person in the class was asked to rate how often they perfomed a number of activities on the Internet on a scale from 0 to 5, where 0 was 'do not use at all', and 5 was 'use extremely often'. The activities included shopping, gaming, Email, Research and Searching, Music Downloads, friends, and photography.

Category III - Computer Experience

Each person in the class answered whether or not they had used various computer technology products in their lives, including Computers, Windows PC's, Macs, IRC, LinuxBox, Excel, Word, OpenOffice, Email, Skype, Wikipedia, Access, and Powerpoint.

Intro – 4. Give Instructions

- Give clear step-by-step Instructions on how to use the spreadsheet

- Especially important if the user needs to later make changes to the data or calculations

Example Intro

24	
25	**Would you like instructions on working with this Excel file?**
26	
27	>>>To view the raw data, and to see calculations done to it, click the 'Data' tab.
28	>>>To view how the ideas for the charts and the formulas were put together, click the 'Model' tab.
29	>>>To better understand any calculations made in this file, click the 'DataDictionary' tab.
30	>>>To view the charts in an Information Dashboard setting, click the 'Dashboard' tab.\
31	

Intro – 5 & 6. Refs and TOCs

- Cite all resources used to create the spreadsheet if any
- Use APA or other format for citation:
  - J. Nevison, *The Elements of Spreadsheet Style*, Prentice-Hall, 1987
- A Table of Contents (TOC) can be a good idea too

Model Section

- Model includes assumptions and calculations

- Tips for a good model section:

1. Explain the model
2. Point to the right source

Model – Explain it

- Should provide three levels of explanation:
  - Explain the values appearing in the model
  - Explain tricky formulas
  - A complete listing of all used formulas

Category I - Most Visited Websites

Calculated Variables: A calculated variable, in the context of this Excel spreadsheet, is any value that is NOT part of the raw data, but has been derived via use of a function or an operator to manipulate a raw data value.

Example: The Total Score was NOT part of the raw data - it was arrived at by counting the number of times each site was the most popular, second most popular, etc., and then multiplying those numbers by scores assigned to them. (See 'Total Score' for details)

Frequency of Websites: The frequency of websites is simply the number of times they showed up in each column (Website1, Website2, Website3, Website4, and Website5). It was calculated through a series of steps:

- Step 1: Using an advanced filter, copy the table with all the websites listed, putting all websites into one column, with any duplicate entries removed.
- Step 2: For the first website in the list, use the COUNTIF(range, criteria) function in Excel to count the number of times it appears in each column listed above.
- Step 3: Fill down for all the other websites in the list. (NOTE: Before you do this, be sure to put '$' in front of the range. For example, in the 'Data' tab, Website1 was cells B3:B30. To use COUNTIF(range, criteria) for Google, one would type in '=COUNTIF(B3:B30,B35), where B35 is the cell labelled "Google". Instead of saying this, in order to fill down properly, you would type in '=COUNTIF($B$3:$B$30,B35). This means that the B3:B30 part will NOT change when you fill down, but the B35 will change every row (B36, B37, etc.)
- Step 4: Use the SUM() function in Excel to add up all of the scores for each of the websites. There should be one sum for each row (website).
- Step 5: Remove the Null category, as it is not a website.
- Step 6: Sort the websites so that they are in order from highest total score to lowest total score, using the Sort Command.

Total Score: The total score for each website was calculated based on a points system, which effectively weighted how popular the site was. The weightings are as follows:

Website	Used...	Points
1	Most	5
2	More	4
3	Average	3
4	Less	2
5	Least	1

Note: All of the values under the column "Used..." are RELATIVE. That is, Website 5 is used "least" in comparison to the other four websites, but not necessarily less than any other website not listed.

The number of times each website showed up in each popularity slot was calculated using a COUNTIF() command (see 'Frequency of Websites'), and then this number was multiplied by the appropriate point value from the above table to give five scores for each website (one for each column - Websites 1-5).

The total score was then determined by summing up the five individual scores.

Use of Conditional Formattin Conditional formatting was used to make it easy to eliminate the "Null" values from the raw data table. Since "Null" does not represent a website, conditional formatting was used to make any cell containing "Null" have a light red background with dark red text. These values were then not included within the Calculated Variables table.

The frequency of websites is simply the number of times they showed up in each column (Website1, Website2, Website3, Website4, and Website5). It was calculated through a series of steps:

Step 1:	Using an advanced filter, copy the table with all the websites listed, putting all websites into one column, with any duplicate entries removed.
Step 2:	For the first website in the list, use the COUNTIF(range, criteria) function in Excel to count the number of times it appears in each column listed above.
Step 3:	Fill down for all the other websites in the list. (NOTE: Before you do this, be sure to put '$' in front of the range. For example, in the 'Data' tab, Website1 was cells B3:B30. To use COUNTIF(range, criteria) for Google, one would type in '=COUNTIF(B3:B30,B35), where B35 is the cell labelled "Google". Instead of saying this, in order to fill down properly, you would type in '=COUNTIF($B$3:$B$30,B35). This means that the B3:B30 part will NOT change when you fill down, but the B35 will change every row (B36, B37, etc.)
Step 4:	Use the SUM() function in Excel to add up all of the scores for each of the websites. There should be one sum for each row (website).
Step 5:	Remove the Null category, as it is not a website.
Step 6:	Sort the websites so that they are in order from highest total score to lowest total score, using the Sort Command.

The total score for each website was calculated based on a points system, which effectively weighted how popular the site was. The weightings are as follows:

Website	Used...	Points
1	Most	5
2	More	4
3	Average	3
4	Less	2
5	Least	1

Note: All of the values under the column "Used..." are RELATIVE. That is, Website 5 is used "least" in comparison to the other four websites, but not necessarily less than any other website not listed.

The number of times each website showed up in each popularity slot was calculated using a COUNTIF() command (see 'Frequency of Websites'), and then this number was multiplied by the appropriate point value from the above table to give five scores for each website (one for each column - Websites 1-5).

The total score was then determined by summing up the five individual scores.

Category II - Internet Interests

Calculated Variables:	See 'Calculated Variables' under 'Category I - Most Visited Websites.
Average (Mean):	This is the sum of all of the scores for that category, divided by the total number of scores. T To enter the average into Excel, the AVERAGE() function was used.
Standard Deviation:	The standard deviation gives an idea of the 'spread' of the data - it gives an idea of how relia To enter the standard deviation into Excel, the STDEV() function was used.
Percentage Scores:	To make the chart simpler to understand, the raw data scores (which were out of 5) were c the original mean score by 5, and then multiplying by 100%.

Category III - Computer Experience

Binary Values:	The binary number system only has two digits, 0 and 1. As a result, any yes/no question ca allows for the conversion of textual data (words) into numeric data that can be statistically a
Calculated Values:	See 'Calculated Variables' under 'Category I - Most Visited Websites.
Percentage Scores:	The percentage scores in this section of the Data tab represent the percentage of students column (since all No answers were 0, and all Yes answers were 1, taking the sum of the co number of students (28), and the result multiplied by 100%.
Use of Conditional Formattin	Conditional formatting was used to separate 1's from 0's - the 1's are all in green.

Data Dictionary

- Explains the meaning of your data
- Give each field:
  1) **Location** (cell range),
  2) **Name**,
  3) The **Data Class** it is (Raw Data, Statistical Summary, Calculated Variable, Score etc.),
  4) **Data Type** (e.g. Integer, Text, Currency, Date, etc.) and
  5) **Description** (a description of the data or what it's 'purpose' is).

Example Data Dictionary

3	**Category I - Most Visited Websites (Data!A1:G77)**			
4	**Name**	**Field Type**	**Data Type**	**Sheet/Cell Reference**
5	Student	Categorization	Text	Data!A3:A30
6	Website(1-5)	Raw	Text	Data!B3:F30
7	Site	Categorization	Text	Data!A35:A77
8	Score (1-5)	Column Calculation	Integer	Data!B35:F77
9	Total Score	Row Calculation	Integer	Data!G35:G77
10				
11	**Category II - Internet Interests (Data!I1:P52)**			
12	**Name**	**Field Type**	**Data Type**	**Sheet/Cell Reference**
13	Student	Categorization	Text	Data!I3:I30
14	Shopping	Raw	Integer	Data!J3:J30
15	Research/Search	Raw	Integer	Data!K3:K30
16	MusicDownloads	Raw	Integer	Data!L3:L30
17	Friends	Raw	Integer	Data!M3:M30
18	Email	Raw	Integer	Data!N3:N30
19	Photography	Raw	Integer	Data!O3:O30
20	Gaming	Raw	Integer	Data!P3:P30
21	Activity	Categorization	Text	Data!I34:I52
22	Mean Score out of 5	Column Calculation	Number	Data!J35:J42
23	Standard Deviation out of 5	Column Calculation	Number	Data!M35:M42
24	Mean Score Percentage	Cell Calculation	Percentage	Data!J45:J52
25	Standard Deviation Percentage	Cell Calculation	Percentage	Data!M45:M52

Example Data Dictionary

3	Category I - Most Visited Websites (Data!A1:G77)			
4	Name	Field Type	Data Type	Sheet/Cell Reference
5	Student	Categorization	Text	Data!A3:A30
6	Website(1-5)	Raw	Text	Data!B3:F30
7	Site	Categorization	Text	Data!A35:A77
8	Score (1-5)	Column Calculation	Integer	Data!B35:F77
9	Total Score	Row Calculation	Integer	Data!G35:G77
10				

a!A1:G77)	
Sheet/Cell Reference	Description
Data!A3:A30	Lists the student number for ease of organization of data.
Data!B3:F30	Lists the 1st favorite, 2nd favorite, 3rd favorite, etc. website of each student.
Data!A35:A77	A filtered list of each website from Data!B3:F30, showing each site only once.
Data!B35:F77	Calculates a Score based on scoring system in "Model" tab, for each column.
Data!G35:G77	Adds up the total score (from columns Score 1 to Score 5); e.g. =SUM(B35:F35)

Data

- Includes raw and calculated data
- This is the actual spread sheet

Example Raw Data (Most Visited Websites)

Most Visited Websites					
Student	Website1	Website2	Website3	Website4	Website5
1	Hotmail	Yahoo	Facebook	Bank	Youtube
2	UofC	Google	Hotmail	Bank	Facebook
3	StockWatch	StockHouse	Kitco	Canucks.com	Arsenal.com
4	UofC	Google	Bank	Null	Null
5	Google	TheLottery	NHL.com	Wikipedia	Funnyjunk.com
6	Yahoo	MySpace	Google	MSN	Null
7	Hotmail	Facebook	Youtube	UofC	MySpace
8	Yahoo	Facebook	Youtube	Hi5	Null
9	Google	Youtube	Wikipedia	NFL.com	Horoscope
10	Google	Youtube	Wikipedia	NHL.com	Hotmail
11	Google	Facebook	Shawlife.com	Calgaryplanet	Hotmail
12	Google	Hotmail	Facebook	Lonelyplanet	Youtube
13	TSN.ca	NHL.com	Google	Gmail	Facebook
14	MSN	Google	Calgaryplus.com	MSNBC.com	Facebook
15	Yahoo	Google	Facebook	Frienndster	Calgary Weather
16	Facebook	Hotmail	Youtube	Google	imdb.com
17	Google	Hotmail	UofC	Facebook	ATB

Example Raw Data (Internet Interests)

Internet Interests							
Student	Shopping	Research/Search	MusicDownloads	Friends	Email	Photography	Gaming
1	0	4	5	3	4	0	0
2	1	4	2	2	4	0	5
3	2	5	4	5	5	1	5
4	3	3	0	4	5	0	0
5	0	4	3	5	5	2	2
6	0	5	0	3	5	0	0
7	0	4	4	5	5	2	0
8	0	3	0	4	5	0	0
9	0	5	5	4	5	5	0
10	0	5	3	4	5	1	0
11	1	5	4	5	3	3	5
12	0	5	4	5	5	1	0
13	4	4	5	5	3	1	2
14	1	4	3	4	4	3	3
15	1	2	4	5	4	4	2
16	0	4	5	5	5	1	3

Example Raw Data

	Computer Experience									
Student	Computer	Email	Word	Excel	OpenOffice	Wikipedia	WindowsPC	Skype	MSPowerpoint	MSAc
1	1	1	1	0	0	1	1	0	0	0
2	1	1	1	1	0	1	1	0	1	1
3	1	1	1	1	1	1	1	0	1	1
4	1	1	1	1	0	1	1	0	1	0
5	1	1	1	1	0	1	1	0	1	0
6	1	1	1	1	0	1	1	0	1	0
7	1	1	1	1	0	1	1	0	1	0
8	1	1	1	0	0	1	1	0	1	0
9	1	1	1	1	0	1	1	0	1	0
10	1	1	1	1	0	1	1	1	1	0
11	1	1	1	1	0	1	1	0	1	0
12	1	1	1	1	0	1	1	1	1	0
13	1	1	1	1	0	1	1	0	1	0
14	1	1	1	1	0	1	1	0	1	0
15	0	0	1	1	0	1	1	0	1	0

Example Calculated Data (Most Visited Websites)

Calculated Variables						
I. Frequency of Each Website in List						
Site	Score 1	Score 2	Score 3	Score 4	Score 5	Total Score
Google	30	32	9	6	1	78
Hotmail	25	20	9	0	2	56
Facebook	20	16	12	4	4	56
Yahoo	25	8	0	0	0	33
UofC	10	0	15	2	0	27
Youtube	0	8	9	2	3	22
MSN	10	4	0	2	1	17
Wikipedia	5	0	6	2	0	13
Gmail	0	8	0	2	0	10
NHL.com	0	4	3	2	0	9
Bank	0	0	3	4	1	8

Example Calculated Data (Internet Interests)

Calculated Variables		
I. Average (Mean) Values and Standard Deviations		
Activity	Mean Score out of 5	Standard Deviation out of 5
Shopping	1.04	1.26
Research/Search	4.25	1.00
MusicDownloads	3.29	1.84
Friends	4.39	0.88
Email	4.61	0.69
Photography	1.61	1.59
Gaming	1.68	1.79
II. Conversion of Mean Values out of 5 to Scores out of 100		
Activity	Mean Score out of 100	Standard Deviation out of 100
Shopping	21	25

Example Calculated Data (Computer Experience)

Calculated Variables										
I. Percentage of Users that have Used Each Program/Computer										
	LinuxBox	IRC	OpenOffice	MS Access	Skype	Mac	Wikipedia	Excel	MSPowerpoint	Emai
Percentage	4%	7%	7%	14%	18%	57%	86%	89%	93%	96

Dashboard

- Visual Charts and conclusions

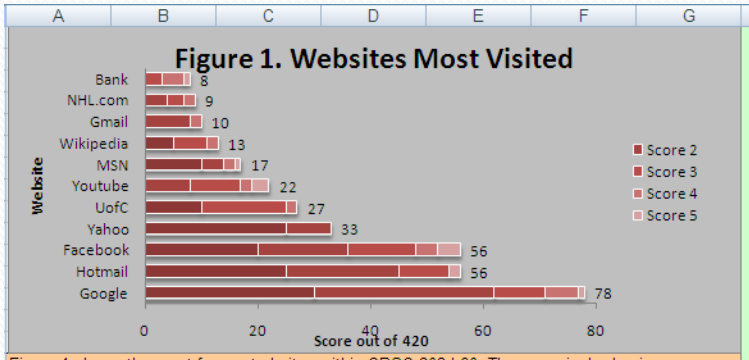


Figure 1 shows the most frequented sites within CPSC 203 L60. The numerical value is a score for that site, calculated from a system with a total of 420 points, where 5 points were assigned to the site if it was the most frequented site by a student, 4 points if it was the second most frequented, etc. Also, each score is coded by a different color in the bar. Sites with less than 8 points were not included in this graph, to allow for clarity.

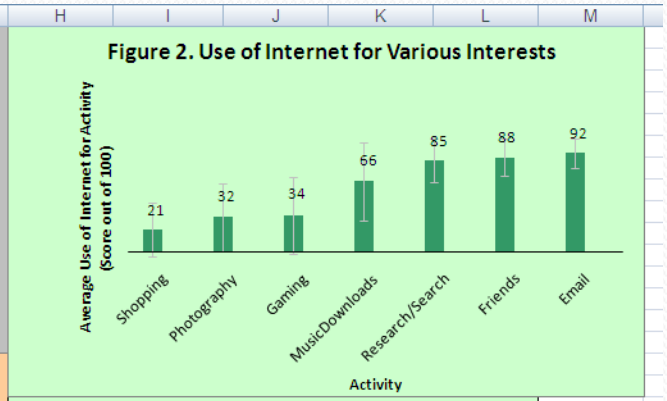


Figure 2 shows how much students rated themselves as using many different activities on the Internet. While the students rated the amount they used each activity out of ascore of 5, this chart shows the score as a percentage. As you can see, shopping had the leastuse, while Email was the most commonly used Internet activity. However, as you can see above, the error bars, which show one standard deviation, are very large, and tend to overlap, showing that there is less confidence in a trend from the data obtained from the class.

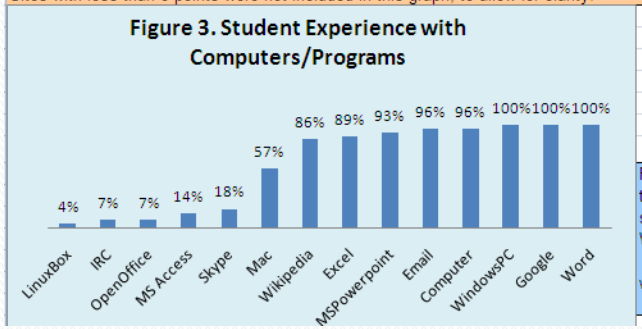


Figure 3 shows the percentage of students that have used the computer hardware or software listed. The chart shows that Windows PCs, Google, and MS Word were the most widely used programs, while LinuxBox, IRC, and OpenOffice were the least commonly used.

Discussion

As shown in these three figures, the students
the Internet the most for friends, searching, a
also reflected in the Most Visited Websites,
Facebook (Friends), and Hotmail (E-mail) wh
visited websites. Also, technologies related t
emailing and friends were the ones with the n
Word (research), Email, Computers, and Win

Example Dashboard

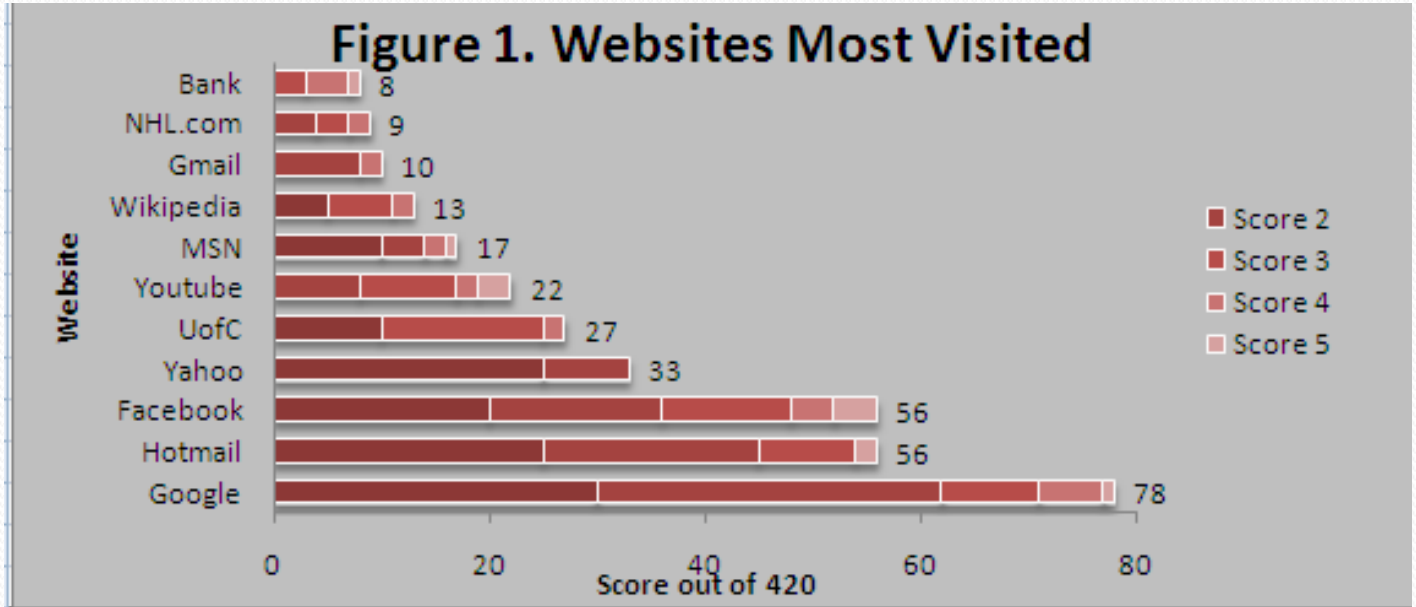


Figure 1 shows the most frequented sites within CPSC 203 L60. The numerical value is a score for that site, calculated from a system with a total of 420 points, where 5 points were assigned to the site if it was the most frequented site by a student, 4 points if it was the second most frequented, etc. Also, each score is coded by a different color in the bar. Sites with less than 8 points were not included in this graph, to allow for clarity.

Example Dashboard

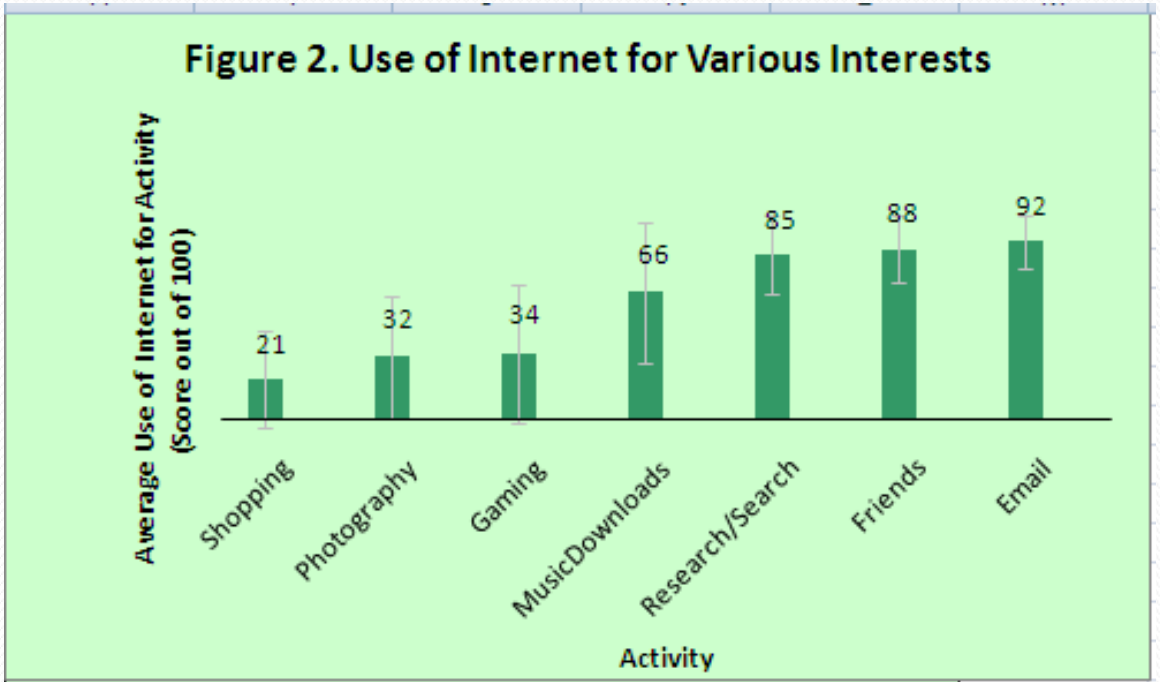
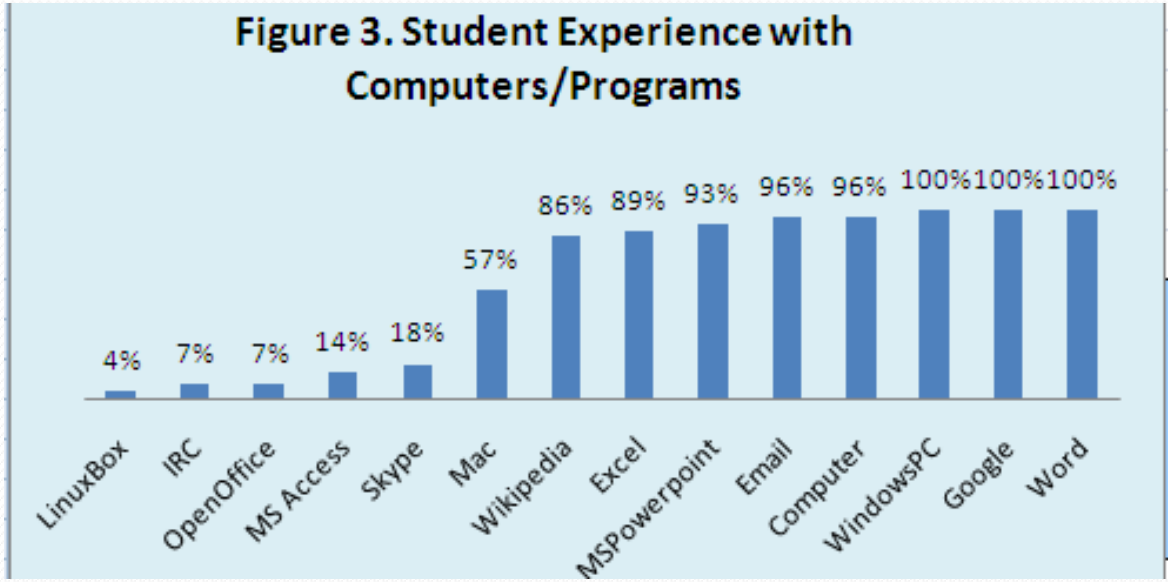


Figure 2 shows how much students rated themselves as using many different activities on the Internet. While the students rated the amount they used each activity out of ascore of 5, this chart shows the score as a percentage. As you can see, shopping had the leastuse, while Email was the most commonly used Internet activity. However, as you can see above, the error bars, which show one standard deviation, are very large, and tend to overlap, showing that there is less confidence in a trend from the data obtained from the class.

Example Dashboard



as a percentage. As you can se
Email was the most commonly
can see above, the error bars, w
are very large, and tend to overla
confidence in a trend from the da

Figure 3 shows the percentage of students that have used the computer hardware or software listed. The chart shows that Windows PCs, Google, and MS Word were the most widely used programs, while LinuxBox, IRC, and OpenOffice were the least commonly used.

Example Dashboard

Discussion

As shown in these three figures, the students in CPSC 203 L60 used the Internet the most for friends, searching, and email. This was also reflected in the Most Visited Websites, where Google (search), Facebook (Friends), and Hotmail (E-mail) where the most commonly visited websites. Also, technologies related to searching, researching, emailing and friends were the ones with the most experience (Google, Word (research), Email, Computers, and Windows PC's)

Overall, the class is fairly computer literate, having experience with a wide range of computers and programs, and having used the Internet quite extensively.

Exercise

- For the Data in the “Data” sheet, make the following
  - Think about some calculations that extract meaningful information from the data
  - Construct a data dictionary sheet for your data
  - Provide at least two graphs in a dashboard sheet
  - Document your graphs
  - Write your conclusion driven from the graph
  - Finally, make the introduction and model sheets

Hints

- Average/total mark for each student
- Average mark for each subject
- Average mark for all subjects
- No. of students in each country
  - Use countif (use Excel help)
- Total marks of students in each country
  - Use sumif (use Excel help)
- Average marks of students in each country