

CS 105

Week 7

Midterm #1

Midterm #1

Expected grades on Wednesday

MP1, MP2

Grade on Compass 2g

Grade Details on CS 105 website

Activity 6, 7

Rethinking activities, making them better; more on this on Wednesday

Data

All data is made up of 0s and 1s.

File extensions tell us how to interpret the 0s and 1s.

File Extensions

- The part of the file **after the dot:**

File Extensions

- The part of the file **after the dot**:

.js: JavaScript code

File Extensions

- The part of the file **after the dot**:

.js: JavaScript code

.html: HyperText Markup Language

File Extensions

- The part of the file **after the dot**:
 - .js**: JavaScript code
 - .html**: HyperText Markup Language
 - .xlsx**: Excel

File Extensions

- The part of the file **after the dot**:
 - .js**: JavaScript code
 - .html**: HyperText Markup Language
 - .xlsx**: Excel
 - .xml**: eXtensible Markup Language

File Extensions

- The part of the file **after the dot**:
 - .js**: JavaScript code
 - .html**: HyperText Markup Language
 - .xlsx**: Excel
 - .xml**: eXtensible Markup Language
 - .csv**: Comma Separated Values

File Extensions

- The part of the file **after the dot**:
 - .js**: JavaScript code
 - .html**: HyperText Markup Language
 - .xlsx**: Excel
 - .xml**: eXtensible Markup Language
 - .csv**: Comma Separated Values
 - .fasta**: DNA Data

Some file types do not make sense
outside of their intended context.

Normal text file	length: 10789 lines: 61	Ln: 18 Col: 1 Sel: 0 0	Dos\Windows	ANSI	INS
------------------	-------------------------	--------------------------	-------------	------	-----

API

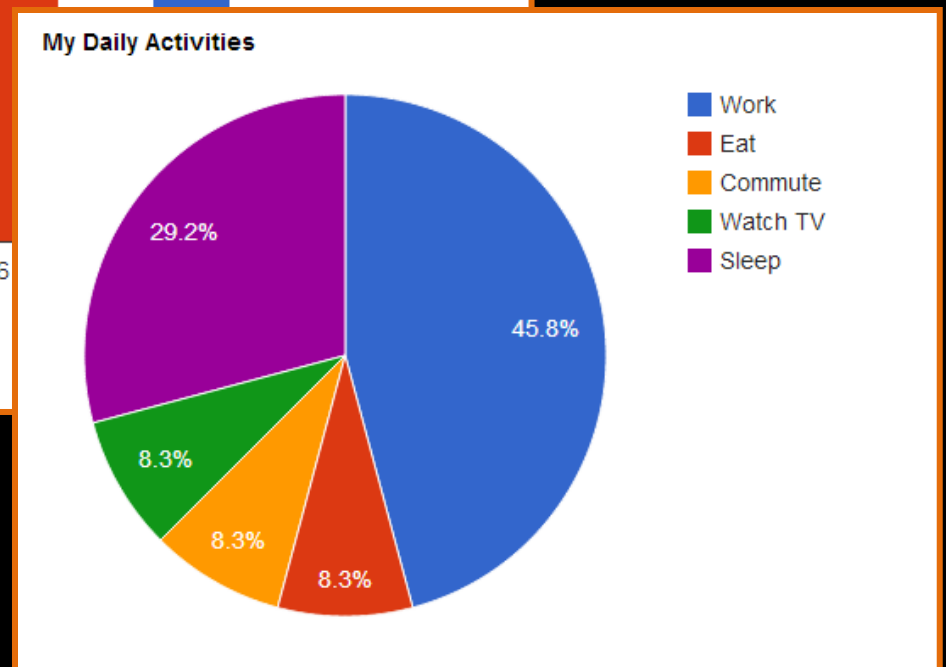
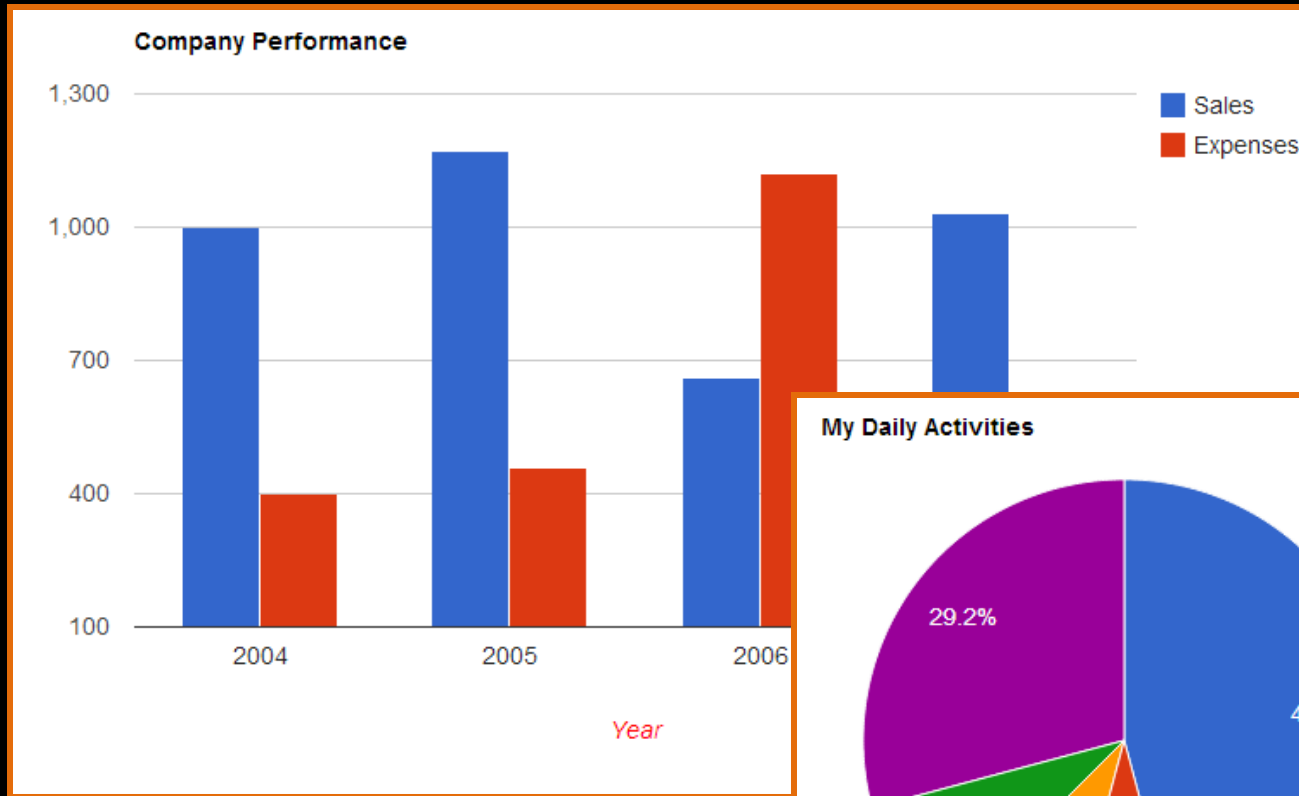
API

Application Programming Interfaces

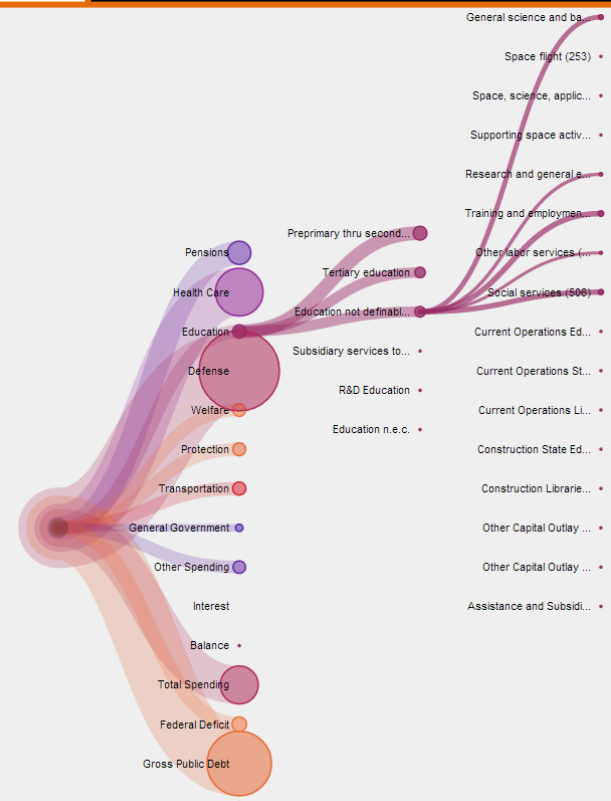
APIs

- Allows a programmer (like you) to access functionality or data from someone else.

Google Graphs API



d3



Data Formats

XML

XML

eXtensible Markup Language

XML

- Very similar to HTML:

XML

- Very similar to HTML:
 - Consists of three types of tags:
start tags: <section>

XML

- Very similar to HTML:
 - Consists of three types of tags:
start tags: `<section>`
end tags: `</section>`

XML

- Very similar to HTML:
 - Consists of three types of tags:
 - start tags: `<section>`
 - end tags: `</section>`
 - empty-element tags: `<section />`

XML

- Very similar to HTML:
 - Consists of three types of tags:
 - start tags: `<section>`
 - end tags: `</section>`
 - empty-element tags: `<section />`
 - All start tags must have an end tag.

XML Example

- Yahoo! Currency API:

<http://finance.yahoo.com/webservice/v1/symbols/allcurrencies/quote>

CSV

CSV

Comma Separated Values

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV Example

- Yahoo! Finance API:

<http://finance.yahoo.com/d/quotes.csv?s=AAPL+GOOG+MSFT&f=nab>

CSV Example

- Yahoo! Finance API:

<http://finance.yahoo.com/d/quotes.csv?s=AAPL+GOOG+MSFT&f=nab>

"Apple Inc.",529.31,528.61

"Google Inc.",1214.90,1214.21

"Microsoft Corpora",37.83,37.78

CSV

- Excel will read CSV files directly!

CSV

- Excel will read CSV files directly!

...so we will often try and get our data in CSV, if it is available.

Application Specific Data

Application Specific Data

- Many applications use their own data formats.

Application Specific Data

- Many applications use their own data formats.

DNA Data uses the **FASTA file format**

Application Specific Data

- Many applications use their own data formats.

DNA Data uses the **FASTA file format**

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEDTREMPEFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLTWETNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIP SANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESQFRADHPFLFLIKHNPTNTIVYFGRYWSP
```

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLTWETNPNTMEKRRVKVYLP
QMKIEEKYNLT SVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPESSEQFRADHPFLFLIKHNPTNTIVYFGRYWSP
```

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLL**VSSST**DLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLT EWTNPNTMEKRRVKVYLP
QMKIEEKYNLT SVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPES EQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLL**VSSST**DLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLTWETNPNTMEKRRVKVYLP
QMKIEEKYNLT SVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPESSEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLT EWTNPNTMEKRRVKVYLP
QMKIEEKYNLT SVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPES EQFRADHPFLFLIKHNPTNTIVYFGRYWSP
```

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLT EWTNPNTMEKRRVKVYLP
QMKIEEKYNLT SVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPESSEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLL**VSSST**DLDTTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDL SMLVLLPDEVSDLERIEKTINFEKLTWETNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIP SANLTGISSAESLKISQAVHGA FMELSEDGIEMAGSTGV
IEDIKHSPESSEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

Excel

Microsoft Excel is a spreadsheet application.

Spreadsheets

Spreadsheets

- Provides users a two-dimensional layout of **cells** that are organized into **rows** and **columns**.

Spreadsheets

- Provides users a two-dimensional layout of **cells** that are organized into **rows** and **columns**.
- Rows are labeled numerically (1, 2, 3...)
- Columns are labeled alphabetically (A, B, C, ...)

Spreadsheets

- Provides users a two-dimensional layout of **cells** that are organized into **rows** and **columns**.
- Excel is one of many spreadsheet tools:
 - Google Docs
 - OpenOffice
 - Numbers for Mac

Microsoft Excel

- Microsoft Excel is the **most widely used** spreadsheet tool.

Data Types

Five Basic Types

Five Basic Types

- Numbers

Five Basic Types

- Numbers
- Strings

Five Basic Types

- Numbers
- Strings
- Dates

Five Basic Types

- Numbers
- Strings
- Dates
- Currencies

Five Basic Types

- Numbers
- Booleans
- Strings
- Dates
- Currencies

Five Basic Types

- Numbers
- Booleans
- Strings
- Dates
- Currencies

Formulas

Excel Formulas

- If the first character in your input is an equal sign (=), **the value of the cell is read as a program.**

Excel Formulas

- If the first character in your input is an equal sign (=), **the value of the cell is read as a program.**
- You can reference other cells' value by entering their cell column and row identification (eg: **=C3**).