

Data ↔ Visualization

From Data To Visualization

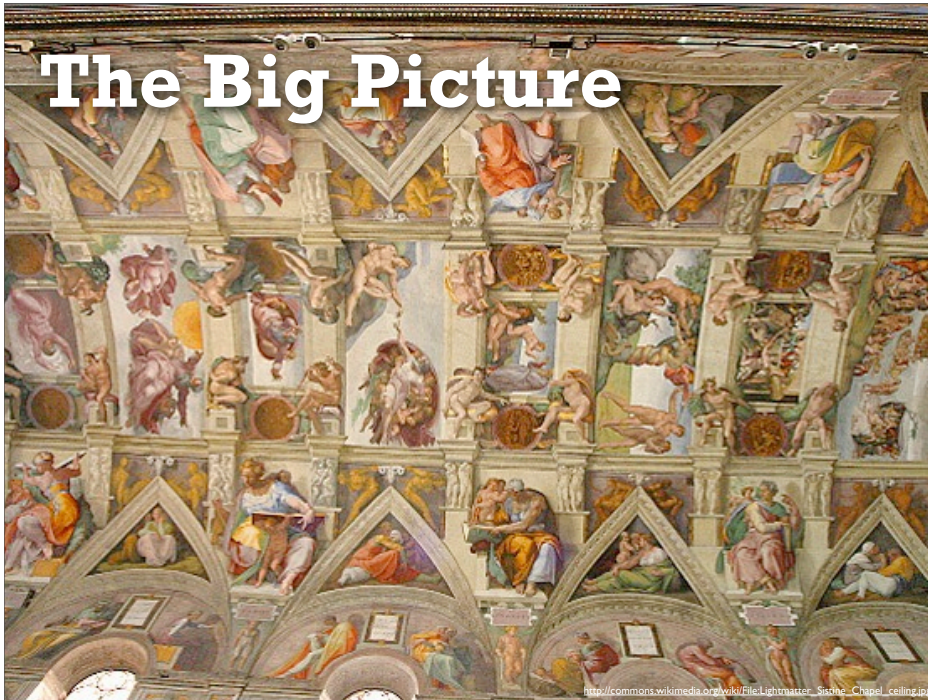
Tuesday, January 24, 12

Last Week

- **History**
- **3 Functions**
 - Record
 - Analyze
 - Communicate
- **Purpose**

Tuesday, January 24, 12

The Big Picture



Tuesday, January 24, 12

The Big Picture

From Data to Image

Data

Abstract type
nominal, ordinal, etc.

Physical type
int, float, etc.

Mapping

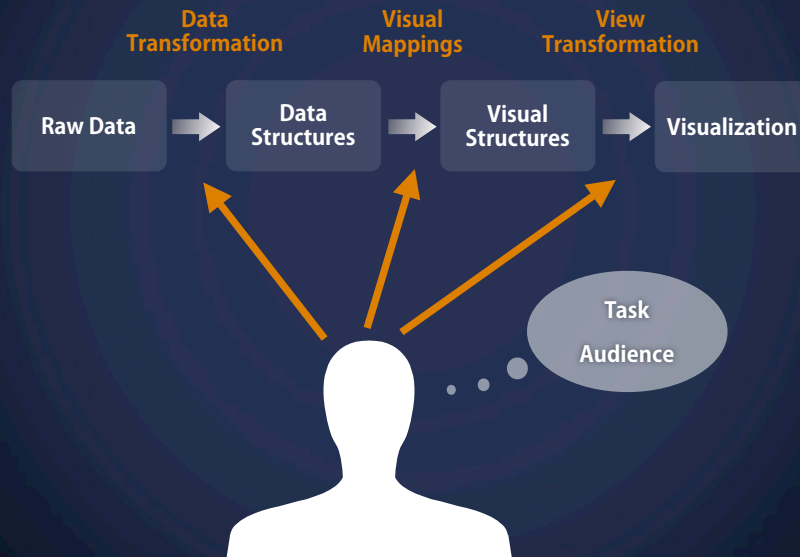
Visual Encoding
Visual Metaphor

Image

Visual channel
Retinal Variables

Tuesday, January 24, 12

The Interaction Model



Tuesday, January 24, 12

Properties of Data

Tuesday, January 24, 12

Statistical Data Models

Month	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	161	163
May	164	158	161	153
June	162	161	158	160
July	166	158	160	148
August	163	158	157	150

Tuesday, January 24, 12

Statistical Data Models

Observations
Cases

Month	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	161	163
May	164	158	161	153
June	162	161	158	160
July	166	158	160	148
August	163	158	157	150

Tuesday, January 24, 12

Statistical Data Models

Categories

Observations
Cases

Month	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	161	163
May	164	158	161	153
June	162	161	158	160
July	166	158	160	148
August	163	158	157	150

Tuesday, January 24, 12

Statistical Data Models

Categories

Variables

Observations
Cases

Month	Control	Placebo	300mg	450mg
March	165	163	166	168
April	162	159	161	163
May	164	158	161	153
June	162	161	158	160
July	166	158	160	148
August	163	158	157	150

Tuesday, January 24, 12

Statistical Data Models

Categories

Variables

Dimensions

Measures

Independent

Dependent

Discrete variables
describing data

Values that can be
aggregated

Tuesday, January 24, 12

Relational Data Models

Timestamp	Department	Student ID	Name	Graduation	Design Skills	Coding Skills	Data Skills	Waitlisted?
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Schema →
Tuples ↘

Timestamp	Department	Student ID	Name	Graduation	Design Skills	Coding Skills	Data Skills	Waitlisted?
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Columns (attributes) have a type

Timestamp	Department	Student ID	Name	Graduation	Design Skills	Coding Skills	Data Skills	Waitlisted?
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Columns (attributes) have a type

Date

Timestamp	Department	Student ID	Name	Graduation	Design Skills	Coding Skills	Data Skills	Waitlisted?
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Columns (attributes) have a type

Date

String

Timestamp	Department	Student ID	Name	Graduation	Design Skills	Coding Skills	Data Skills	Waitlisted?
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Columns (attributes) have a type

Date String Number

Timestamp	Department	Student ID	Name	Graduation	Design	Coding	Data	Waitlisted?
					Skills	Skills	Skills	
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

Relational Data Models

Columns (attributes) have a type

Date String Number Binary

Timestamp	Department	Student ID	Name	Graduation	Design	Coding	Data	Waitlisted?
					Skills	Skills	Skills	
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

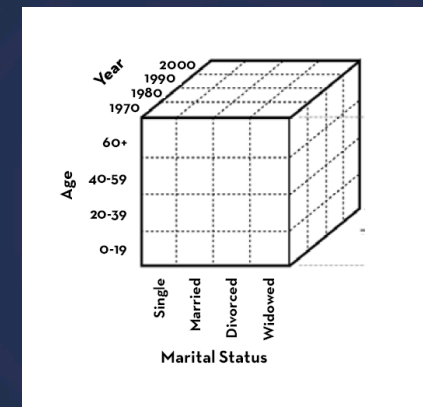
Tuesday, January 24, 12

Cube



Tuesday, January 24, 12

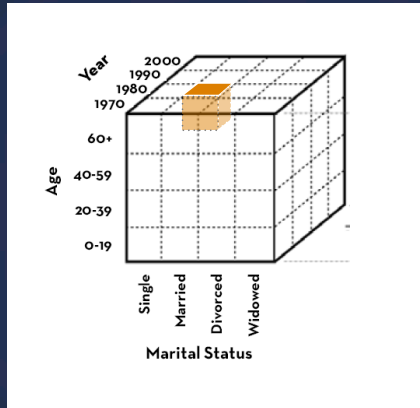
Cube



Adapted from Heer

Tuesday, January 24, 12

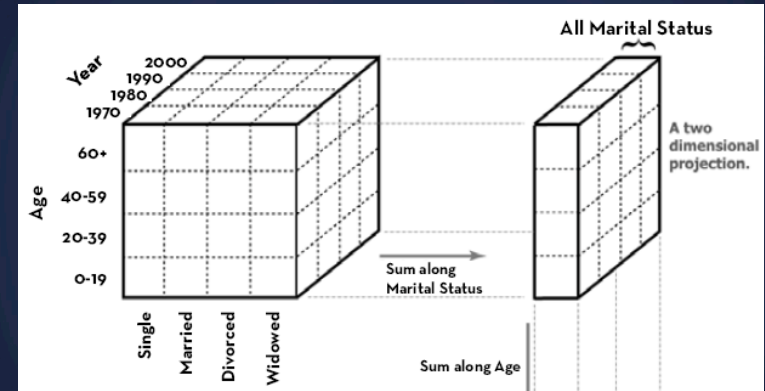
Cube



Adapted from Heer

Tuesday, January 24, 12

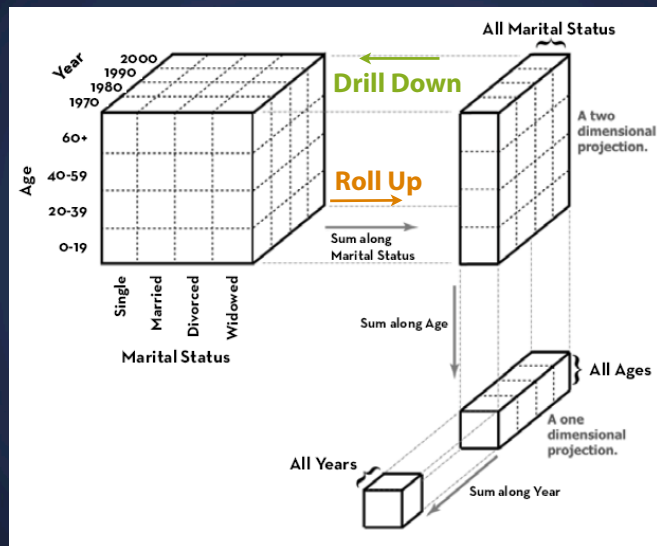
Cube



Adapted from Heer

Tuesday, January 24, 12

Cube



Adapted from Heer

Tuesday, January 24, 12

Cube - Rolling Up

Grouping along desired dimension

```
SELECT year, marital_status, SUM(people)
FROM census
GROUP BY year, marital_status
```

Tuesday, January 24, 12

Taxonomy: Data Types

- 1D (sets, sequences, text)
- 2D (maps)
- 3D (shapes)
- Temporal
- Multidimensional (relations)
- Tree (hierarchies)
- Network (graphs)

Each data type allows for a specific set of tasks

Shneiderman: "The Eyes Have It"

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)



Ordinal

ordered



A, AA, AAA

1st, 2nd, 3rd

Quantitative

Interval

Gap comparison

lat/long, °F

Ratio

Size comparison
(0 is fixed)

inches

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)

Ordinal

ordered

Quantitative

Interval

Gap comparison

Ratio

Size comparison

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)

= ≠

Ordinal

ordered

= ≠

Quantitative

Interval

Gap comparison

Ratio

Size comparison

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)

= ≠

Ordinal

ordered

= ≠
< > ≤ ≥

Quantitative

= ≠ < > ≤ ≥

Interval

Gap comparison

—

Ratio

Size comparison

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)

= ≠

Ordinal

ordered

= ≠
< > ≤ ≥

Quantitative

= ≠ < > ≤ ≥

Interval

Gap comparison

—

Ratio

Size comparison

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Taxonomy: Attribute Quality

Nominal

Labels (unordered)

= ≠

Ordinal

ordered

= ≠
< > ≤ ≥

Quantitative

= ≠ < > ≤ ≥

Interval

Gap comparison

—

Ratio

Size comparison

÷

S. S. Stevens, On the theory of scales of measurements, 1946

Tuesday, January 24, 12

Transformations

Nominal

Ignore ordering

Sort
e.g. alphabetical

Ordinal

Movie length to
short, medium, long

Tuesday, January 24, 12

Example

Temperature (Conceptual Model)

Data Model

°F : 32.5, 54.0, -17.3

°C : 0.2, 12.2, -27.3

Nominal

Burned
vs
Not burned

Ordinal

Hot
Warm
Cold

Quantitative

Continuous range
of values

based on slide from Munzner

Tuesday, January 24, 12

Example

Class Survey: Your turn.

Timestamp	Department	Student ID	Name	Graduation	Design	Coding	Data	Waitlisted?
					Skills	Skills	Skills	
Jan 13, 2012 12:11	School of Information	35489234	Michael Palin	05/2014	4	2	3	No
Jan 13, 2012 17:00	EECS	23456325	John Cleese		3	5	4	Yes
Jan 13, 2012 18:40	School of Information	10324324	Terry Gilliam	05/2012	2	4	1	No
Jan 14, 2012 11:08	Haas Business School	3546424	Terry Jones	05/2013	3	3	4	No
Jan 15, 2012 10:06	School of Information	56589321	Eric Idle	05/2013	4	2	2	No
Jan 15, 2012 10:17	School of Information	78962218	Carol Cleveland	05/2012	5	3	3	No
Jan 17, 2012 17:04	Psychology	95185633	Graham Chapman	12/2015	1	3	5	Yes

Tuesday, January 24, 12

The Big Picture

From Data to Image

Data

Abstract type
nominal, ordinal, etc.

Physical type
int, float, etc.

Mapping

Visual Encoding
Visual Metaphor

Image

Visual channel
Retinal Variables

Tuesday, January 24, 12

Properties of The Image

Tuesday, January 24, 12

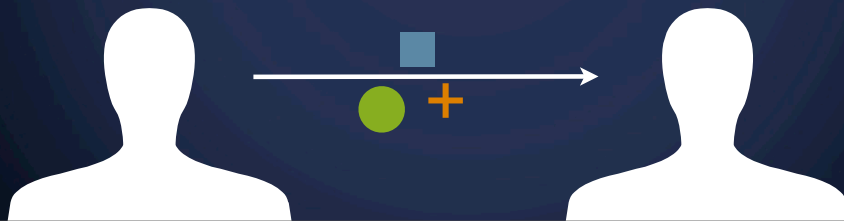
Bertin's Semiology of Data

Visual language is a sign system

- Images perceived as a set of signs
- Sender encodes information in signs
- Receiver decodes information from signs

Sender

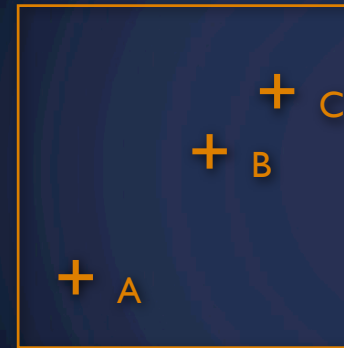
Receiver



Tuesday, January 24, 12

Bertin's Semiology of Data

Information in the position



- A, B, and C are distinguishable
- B is between A and C
- AB is twice as long as BC

"Resemblance, order and proportion are the three signfields in graphics." - Bertin

Tuesday, January 24, 12

Bertin's Semiology of Data

Visual Variable Encodings

LES VARIABLES DE L'IMAGE			
	POINTS	LIGNES	ZONES
XY 2 DIMENSIONS DU PLAN			
Z TAILLE			
VALEUR			
LES VARIABLES DE SÉPARATION DES IMAGES			
GRAIN			
COULEUR			
ORIENTATION			
FORME			

[Bertin, Semiology of Graphics, 1983]

Tuesday, January 24, 12

Bertin's Semiology of Data

Visual Variable Encodings

- Position (2D)
- Size
- Value
- Texture
- Color
- Orientation
- Shape

Also

- Movement (Time)
- 3D
- Transparency
- Blur
- etc.

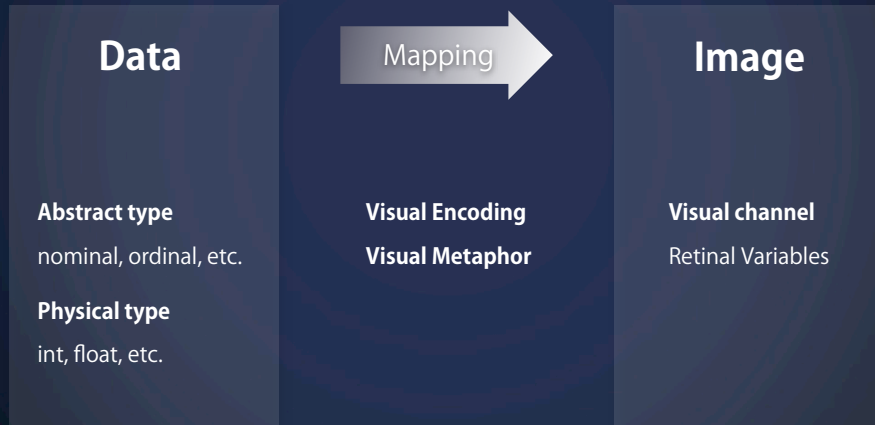
LES VARIABLES DE L'IMAGE			
	POINTS	LIGNES	ZONES
XY 2 DIMENSIONS DU PLAN			
Z TAILLE			
VALEUR			
LES VARIABLES DE SÉPARATION DES IMAGES			
GRAIN			
COULEUR			
ORIENTATION			
FORME			

[Bertin, Semiology of Graphics, 1983]

Tuesday, January 24, 12

The Big Picture

From Data to Image



Tuesday, January 24, 12

From Data to Image

Mapping and (not so much) Magic

Tuesday, January 24, 12

Mappings

Information in size, color, and value

Categorical Nominal (in some cases)
Ordered Ordinal
Continuous Quantitative

Tuesday, January 24, 12

Mappings

Information in size, color, and value

	Size	
Categorical Nominal (in some cases)		
Ordered Ordinal	● ● ● ● ● ●	
Continuous Quantitative	● ● ● ● ● ●	

Tuesday, January 24, 12

Mappings

Information in size, color, and value

	Size	Color	
Categorical Nominal	(in some cases)		
Ordered Ordinal			
Continuous Quantitative			

Tuesday, January 24, 12

Mappings

Information in size, color, and value

	Size	Color	Value
Categorical Nominal	(in some cases) 		
Ordered Ordinal			
Continuous Quantitative			

Tuesday, January 24, 12

Bertin's Semiology of Data

Levels of Organization

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	o	
Color	N		
Orientation	N		
Shape	N		

- N = Nominal
- O = Ordinal
- Q = Quantitative

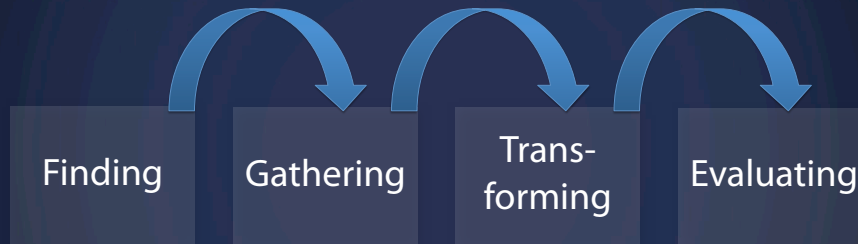
Tuesday, January 24, 12

Process

Finding. Gathering. Transforming.

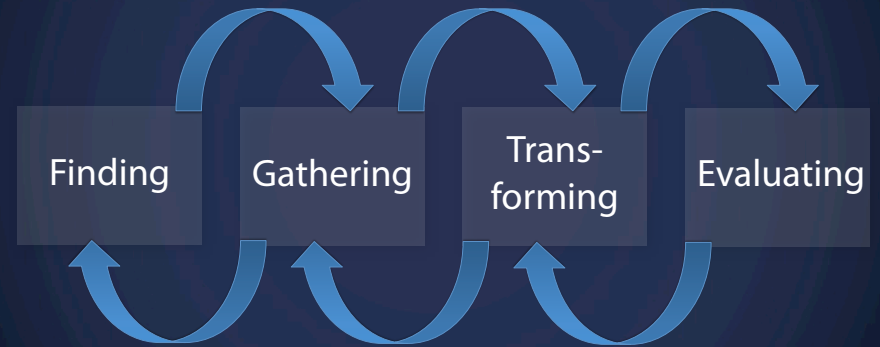
Tuesday, January 24, 12

Process



Tuesday, January 24, 12

Process



Tuesday, January 24, 12

Finding data

The needle in the haystack

Open Government Data



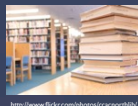
Product APIs



Third Party Data Providers



Anywhere



<http://www.flickr.com/photos/ccacore800/4118279460/size/orig/>

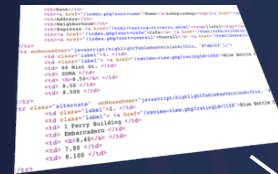
Week	Hour	Count	Rating
1	0	1	0
2	0	2	0
3	0	3	0
4	0	4	0
5	0	5	0
6	0	6	0
7	0	7	0
8	0	8	0
9	0	9	0

Tuesday, January 24, 12

Gathering data

The tedious part

Scraping



Collecting manually

Crowdsourcing



B	C	D	E	F	G	H
link	name	address	neighborhood	rating	overall id	rating_espreso
Review-view	Blue Bottle	C 66 Mint St.	SOMA	8.5	1064	8.5
Review-view	Blue Bottle	C 1 Ferry Buil	Embarcadero	8.1	1128	8.4
Review-view	Blue Bottle	C 315 Linden	S Hayes Valley	8.3	820	8.4
Review-view	Coffee Bar	1890 Bryant	Potrero Hill	8.45	1059	8.4
Review-view	Epicenter	Ca 764 Harrison	SOMA	8.3	1121	8.4
Review-view	Four Barrel	C 375 Valencia	Mission	8.45	1070	8.4
Review-view	Mercury Caf	201 Octavia	Hayes Valley	8.2	1171	8.4
Review-view	Ritual Coffee	1634 Jerrold	Bayview	8.2	1016	8.4
Review-view	Ritual Coffee	1026 Valenc	Mission	8.45	843	8.4
Review-view	Cafe Capricci	2200 Mason	North Beach	8.05	1127	8.3
Review-view	Sendak Bros	24 4th Ave	SOMA	9.05	1122	8.1

Tuesday, January 24, 12

Transforming and evaluating data

Clean your data

- Missing data points
 - inconsistent formats
- Jan 24, 2012
2012/01/24
24/01/12
24th January 2012

Tuesday, January 24, 12

Transforming and evaluating data

Clean your data

- Missing data points
 - inconsistent formats
- Jan 24, 2012
2012/01/24
24/01/12
24th January 2012

Identify your target format

SQL Database?
CSV?
Excel?
JSON? API?

Tuesday, January 24, 12

Recap

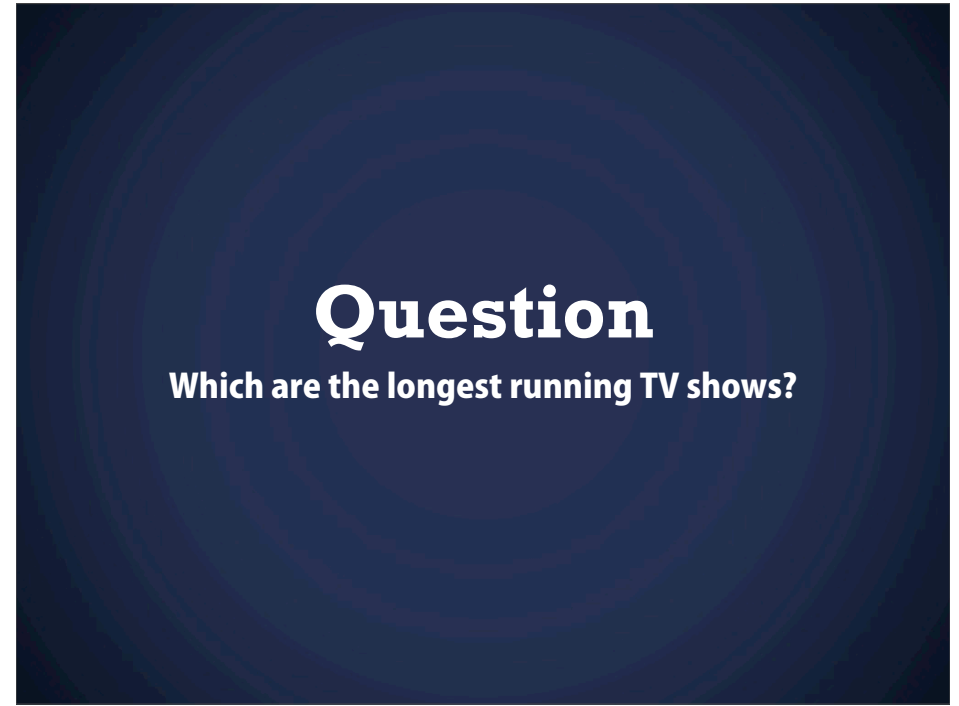
Tuesday, January 24, 12

Lab Data

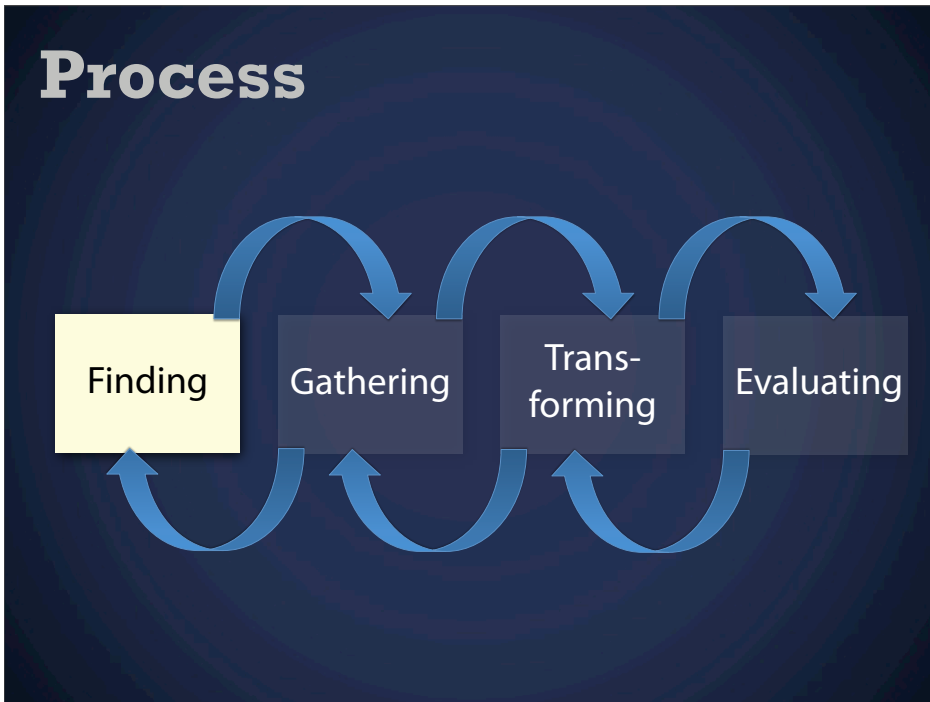
Tuesday, January 24, 12



Tuesday, January 24, 12



Tuesday, January 24, 12



Tuesday, January 24, 12

WIKIPEDIA
The Free Encyclopedia

Log in / create account

Article Talk

List of television programs by episode count

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help **improve this article** by adding citations to **reliable sources**.
Unourced material may be **challenged** and **removed**. (January 2009)

This is a list of **episodic television programs by episode count** with 150 episodes minimum. Episode numbers for ongoing daytime dramas, such as soap operas, are drawn from the websites for the shows. Daily news broadcasts, such as *The Today Show* and *Good Morning America*, are not episodic in nature and are not listed.

Contents [hide]

- Television programs
- See also
- Notes
- External links

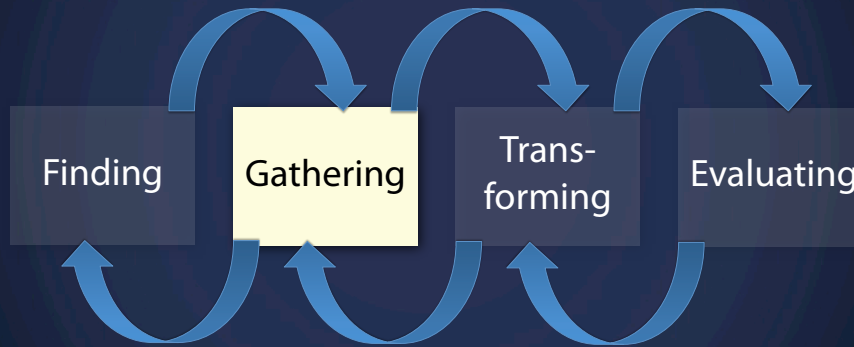
Television programs [edit]

This section **needs additional citations for verification**. Please help **improve this article** by adding citations to **reliable sources**.
Unourced material may be **challenged** and **removed**. (November 2011)

Program Name	Genre	Country of Origin	Episode Length (minutes)	Seasons	Started Broadcasting	Finished Broadcasting	Episode Count
<i>Guiding Light</i>	Soap opera	 United States	60, 30, 15	57	June 30, 1952	September 18, 2009	15,782 ^[1]
<i>As the World Turns</i>	Soap opera	 United States	60, 30	54	April 2, 1956	September 17, 2010	13,858 ^[2]
<i>General Hospital</i>	Soap opera	 United States	60, 45, 30	—	April 1, 1963	Currently in production	12,401 ^[3]
<i>Days of our Lives</i>	Soap opera	 United States	60, 30	—	November 8, 1965	Currently in production	11,679 ^[4]
<i>One Life to Live</i>	Soap opera	 United States	60, 45, 30	43	July 15, 1968	January 13, 2012	11,096 ^[5]

Tuesday, January 24, 12

Process



Tuesday, January 24, 12

Tip

HTML pages to Excel

Google Docs

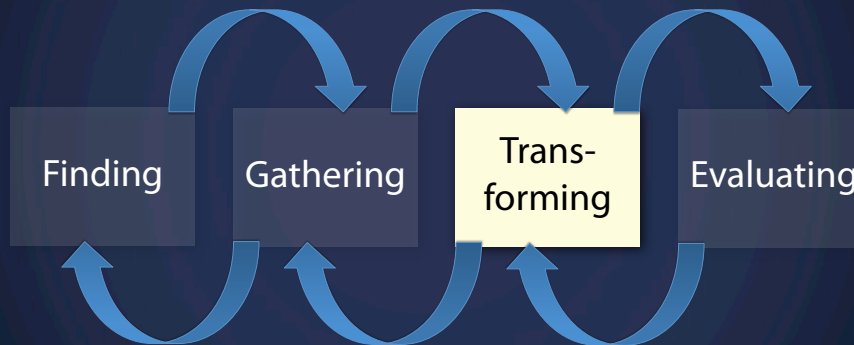
```
=importHtml(URL, element, index)
```

```
=importHtml("http://en.wikipedia.org/wiki/  
List_of_television_programs_by_episode_count", "table", 4)
```

<http://googlesystem.blogspot.com/2007/09/google-spreadsheets-lets-you-import.html>

Tuesday, January 24, 12

Process



Tuesday, January 24, 12

Data Cleaning

Stanford Text Wrangler

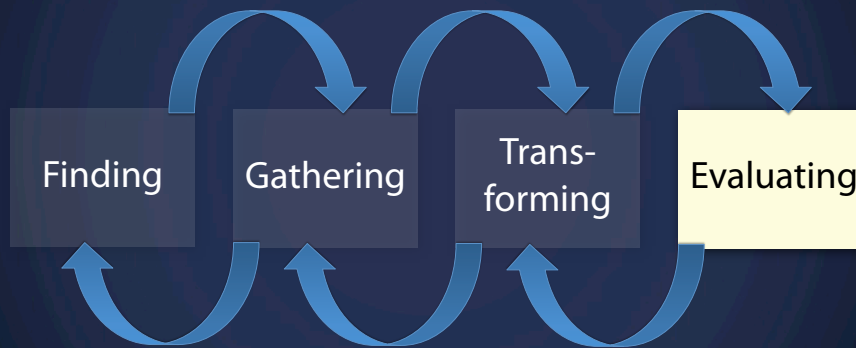
<http://vis.stanford.edu/wrangler/app/>

Google Refine

<http://code.google.com/p/google-refine/wiki/Downloads>

Tuesday, January 24, 12

Process



Tuesday, January 24, 12

Announcement

Assignment 1

Task Create a static visualization from a provided dataset

Dataset Longest Running TV series

Deliverable visualization; 1 page writeup

Due Tuesday Jan 31, 5:00PM

- Find and visualize something interesting from the dataset
- Tell a story
- You can use any tool you like

Dataset and more information on the class blog

Tuesday, January 24, 12