600.315 Demo Presentation

Joon Hyuck Choi (jchoi100), Joo Chang Lee (jlee381)

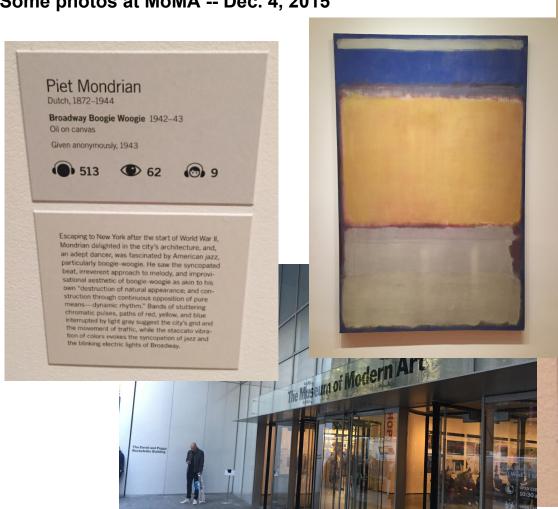
Contents

- 1. Overview / Domain
- 2. Difficulties Faced (4)
- 3. Stored Procedures
- 4. Advanced Topics / What We Have Learned
- 5. Future Development

1. Overview / Domain

- MoMA Database
 - a. Source: https://github.com/MuseumofModernArt/collection
- War Database
 - a. Source: http://www.correlatesofwar.org/
- Demonym Table: wikipedia
- Some photos...

Some photos at MoMA -- Dec. 4, 2015





Marc Chagall

French, born Belarus. 1887-1985

I and the Village 1911

Oil on canvas

Mrs. Simon Guggenheim Fund, 1945

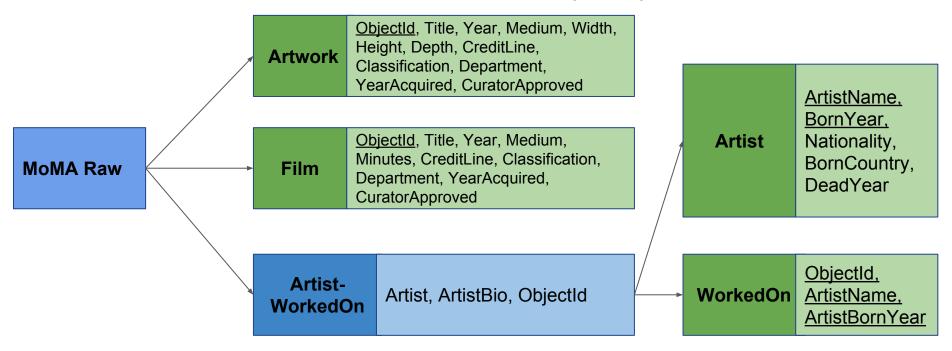




1. MoMA Database (Raw artwork.csv)

2011	I I I I I I I I I I I I I I I I I I I	F - 90 1 - 9.54		U SECTI	M. 11 1862 11	19090.54.1	4	NA CHARLES CONTROL		Curator	£ an
Title	Artist	ArtistBio	Date	Medium	Dimensions	CreditLine	MoMANumber	Classification	Department	DateAcquirec Approv	ObjectID URL
1										ed	1
The Starry Night	Vincent van Gogh	(Dutch, 1853-1890)	Saint Remy, June 1889	Oil on canvas	29 x 36 1/4" (73.7 x 92.1 cm)	Acquired through the Lillie P. Bliss Bequest	st 472.1941 I	Painting	Painting & Sculpture	1941-06-10 Y	79802 http:
The Red Studio	Henri Matisse	(French, 1869-1954)	Issy-les-Moulineaux, fall 1911	1 Oil on canvas	71 1/4" x 7' 2 1/4" (181 x 219.1 cm)	, Mrs. Simon Guggenheim Fund	8.1949	Painting	Painting & Sculpture	1949-02-03 Y	78389 http:
Villa dall'Ava, Pari	is Rem Koolhaas, Madelon Vriesendorp	(Dutch, born 1944) (Dutch, born 1945)	198	7 Synthetic polymer p	oa 25 x 39" (63.5 x 99.1 cm)	Gift of Max Underwood	21.2	A&D Architectural Drawing	Architecture & Design	2000-01-19 Y	82 http
4							+				4

1. MoMA Database: How we broke the tables up and parsed the information



2. War data table (Raw Inter-StateWarData.csv, Extra-StateWarData.csv, Intra-

StateWarData, Non-StateWarData)

StateNan Side

230 Spain

220 France

640 Ottoman

1 Franco-Sp

1 Franco-Sp

4 First Russo

Ext	a-Sta	teWarD	ata_v4	.0.csv																								
4	Α	В	С	D	E	F	G	Н	- 1	J	K	L	M	N	0	Р	Q	R	S	T	U	V	W	X	Υ	Z	AA	AB
1	VarNun	n WarNan	WarTyp	e ccode	1 SideA	ccode2	SideB	StartMon	StartDay*	StartYear	EndMontl E	ndDay1	EndYear1 S	tartMon St	tartDay S	tartYear E	ndMontl E	ndDay2	EndYear2 I	nitiator I	nterven	FransFror C	Outcome	TransTo	WhereFor B	atDeath	NonState V	ersion
2	30	O Allied Bo	n	3	210 Netherlan	-8	3 -8	8	26	1816	8	30	1816	-8	-8	-8	-8	-8	-8	1	1	-8	1	-8	6	13	-8	4
3	30	O Allied Bo	n	3	200 United Kir	-8	3 Algeria	8	26	1816	8	30	1816	-8	-8	-8	-8	-8	-8	1	1	-8	1	-8	6	129	6000	4
4	30	1 Ottoman	-1	3	640 Ottoman	-8	Saudi Wal	9	-9	1816	9	11	1818	-8	-8	-8	-8	-8	-8	1	0	-8	- 1	-8	6	13500	14000	4
Ir	ntra-Si	tateWa	Data_v		D E	F	G	Н	1	J	K	L	M	N	0	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	WarN	lum WarN	am WarT	ype Ccc	odeA SideA	Ccode	B SideB	Intnl	StartMo	on StartDa	y StartYea	er EndMo	ntl EndDay1	EndYear1	StartMo	n StartDay	StartYea	r EndMo	ntl EndDay2	EndYear	2 TransFr	or WhereFo	Initiator	Outcom	e TransTo	SideADe	ea SideBDe	Version
2		500 First C	auc	5	365 Russia		-8 Georgian	ns	0	6	10 181	8	-9 -9	1822	-8	-8	3 -	3	-8 -8	3 -8	3 -	8 2	2 Checher	ns	1 -8	500	0 6000	4.
3		501 Sidon	-Dar	6	-8 Sidon		-8 Damasc	us	0	6	-9 182	0	7 21	1821	-8	-8	3 -	3	-8 -8	3 -8	3 -	8 6	Sidon		2 -8		9 -9	4.
4		502 First T	wo :	4	300 Austria		-8 -	-8	1	3	-9 182	1	3 23	1821	-8	-8	3 -4	3	-8 -8	3 -6	3 -	8 2	2 Liberals		1 -8	-	9 -8	4.
N	on-Sta	ateWarl	Data_v4		/ D E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R	S	Т	U	V	W	X	Υ	Z	AA	AB
1	WarNu	ım WarNa	m WarTy	pe Whe	reFot SideA1	SideA2	SideB1	SideB2	SideB3	SideB4	SideB5	StartYea	ar StartMon	StartDay	EndYear	EndMont	EndDay	Initiator	TransFron	TransTo	Outcome	SideADea	SideBDe	TotalCon	Version			
2	15	00 First M	or	8	9 Te Raup	ar	-8 Taranaki	Ngai Tah	u Waikato	Ngati Ira	Rangitikei	181	8 -9	-9	1824	-9	-9	A	-8	-8	1	1500	6000	7500	4			
3	15	01 Shaka 2	Zul	8	4 Shaka Z	ul	-8 Bantu	1	8 -8	3 -4	8 -8	181	9 -9	-9	1828	9	24	A	-8	-8	1	20000	40000	60000	4			
X	Inte	r-StateV	VarDat	=).csv								M	KI.	0	D		R	6	Ŧ	U	V	102	V	V	7		A.D.
1	A	D	-		D E	F .	G	H		1		L	M	N	0	7	Q	I.	2		0	V	W	X	Y	7	AA	AB

StartMon StartDay StartYear EndMontl EndDay1 EndYear1 StartMon StartDay StartYear EndMontl EndDay2 EndYear2 TransFror WhereFoi Initiator

Non-State WarNum, WarName, WarType,
WhereFought, SideA1, SideA2, SideB1, ...
, StartYear, ..., EndYear, Initiator,
TransFrom, TransTo, Outcome,
SideADeath, SideBDeath,
TotalCombatDeaths

Intra-State WarNum, WarName, WarType, CcodeA, SideA, CcodeB, SideB, Intl, StartMonth1, ..., EndYear2, TransFrom, WhereFought, Initiator, Outcome, TransTo, SideADeath, SideBDeath

Extra-State WarNum, WarName, WarType, ccode1, SideA, ccod2, SideB, StartMonth1, StartDay1, ..., EndYear2, Initiator, Interven, TransFrom, Outcome, TransTo, WhereFought, BatDeath, NonState

Inter-State WarNum, WarName, WarType, ccode, StateName, Side, StartMonth, StartDay, StartYear, EndMonth1, ..., EndYear2, TransFrom, WhereFought, Initiator, Outcome, TransTo, BatDeath

WarData-Combined WarNum, WarName, WarType, ccode, StateName, Side, StartYear, EndYear, Initiator, Outcome, WhereFought, BatDeath Particip ated-In

WarNum, ccode, StateName, Side, Initiator, Outcome, BatDeath

Wars

WarNum,
WarName,
WarType,
StartYear,
EndYear,
WhereFought

2. Difficulties Faced (1)

1. Parsing tables with messy/irregular data

- a. various java programs written:
 - i. ArtistTableMaker.java, ArtworkParser.java, FilmParser.java,
 FilmSeparator.java, HistoryParser.java, MomaArtistParser.java,
 WorkOnTableMaker.java, etc.

2. Difficulties Faced (2)

2. Irregularity of data in the raw MoMA table

- a. ex) Year column in Artwork table
 - some examples: c. 1986, c.1986, Unkown, After 1986, Before 1986, early 1980s, 11/14/1986, newspaper published. 1986, 1986-87, 1986
- b. ex) ArtistBio column in Artist table
 - . some examples: (French and Swiss, born Switzerland 1944), (Dutch, born 1944) (Dutch, born 1945), (American, born Estonia. 1901-1974)

2. Difficulties Faced (2 continued)

- Irregularity of data in the raw MoMA table (continued)
 - c. ex) Dimensions column
 - i. some examples: 14 x 18" (35.6 x 45.7 cm), 28 x 34 3/8 x 34 3/8",

(71.1 x 87.3 x 87.3 cm), I. 16' 2" x w. 31 1/2" (I. 492.8 x w. 80 cm)

2. Difficulties Faced (3)

3. Combining (unioning) tables with incompatible domain types

1816

1816

a. ex)

200 United Kir

640 Ottoman I

-8 Algeria

-8 Saudi Wal

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA AB

WarNum WarNam WarType ccode1

SideA ccode2 SideB StartMon StartDay: StartYear EndMontI EndDay2 EndYear1 StartWon StartDay: StartYear EndMontI EndDay2 EndYear2 Initiator Interven TransFror Outcome TransTo WhereFor BatDeath NonState Version

2 300 Allied Bon 3 210 Netherlan -8 -8 8 26 1816 8 30 1816 -8 -8 -8 -8 -8 -8 -8 1 1 -8 6 13 -8 4

129

6 13500

6000

14000

1816

1818

Intra-StateWarData v4.1.csv

Extra-StateWarData v4.0.csv

300 Allied Bon

301 Ottoman-

- 18	A	В	СГ	D E	F	G	Н	1	J	K	L	M		N	0	Р	Q	R	S	T	U	VVV	/ X	Y	Z	AA	AB
1	WarNum	WarNam(WarType Ccode	eA SideA	CcodeB	SideB	Intnl	StartMo	n StartDa	y StartYe	ear EndM	Iontl EndD	ay1 Enc	dYear1 5	StartMon	StartDay	StartYear	r EndMor	ntl EndDay	2 EndYear	2 TransFro	r WhereFot Initia	tor Outco	me TransT	o SideAD	ea SideBDe	a Version
2	500	First Cauc	5	365 Russia	7	-8 Georgians	s 7	0 6	6 1	10 181	18	-9	-9	1822	-8	-8	-8	1 -	8 -	8 -8	8 -8	3 2 Chec	hens	1	-8 500	00 6000	3 4.1
3	501	Sidon-Dar	6	-8 Sidon	-/	-8 Damascus	AS /	0 /	6 -	9 187	20	7	21	1821	-8	-8	-8	1	- 8	-8	3 -8	6 Sidor	1	2	-8	-9 -1	9 4.1
4	502	2 First Two :	4	300 Austria	-5	8 -8	3	1 /	3 -	-9 182	21	3	23	1821	-8	-8	-8	ş =	8 -	8 -8	3 -8	3 2 Libera	als	1	-8	-9 -1	3 4.1

Non-StateWarData_v4.0.csv

1	A	В	С	D	Е	F	G	Н	1	J	K	L	M	N	0	P	Q	R	S	I	U	V	W	X	Y	Z	AA	AB
1	WarNun	m WarNam W	arType	WhereFot Si	deA1	SideA2	SideB1	SideB2	SideB3	SideB4	SideB5	StartYear	StartMon S	tartDay E	EndYear	EndMont	EndDay	Initiator	TransFror T	FransTo	Outcome	SideADea S	sideBDea 1	TotalCom Vers	ion			
2	1500	0 First Maor	8	9 Te	Raupar		8 Taranaki	Ngai Tahi	u Waikato	Ngati Ira	Rangitikei	1818	-9	-9	1824	-9	-9	A	-8	-8	1	1500	6000	7500	4			
3	150	1 Shaka Zul	8	4 Sh	naka Zul	1.0	8 Bantu	-8	-8	-8	-8	1819	-9	-9	1828	9	24	1 A	-8	-8	- 1	20000	40000	60000	4			

Inter-StateWarData v4.0.csv

1	A	В	С	D	E	F	G	Н	1	J	K	L	M	N	0	P	Q	R	S	T	U	V	W	X	Υ	Z	AA	AB
1	WarNum	WarNam	WarType	ccode	StateNan	Side	StartMo	n StartDay	StartYear	EndMontl	EndDay1	EndYear1	StartMor	StartDay	StartYear	EndMont	EndDay2	2 EndYear	TransFro	WhereFor	Initiator	Outcome	TransTo	BatDeat	Version			
2	1	Franco-Sp	1	23	30 Spain	1	2 4	7	1823	11	13	1823	-8	-8	-8	-8	-8	8 -8	503	2	2	2	9	8 600	4			
3	1	Franco-Sp	1	22	20 France	S.	1 4	7	1823	11	13	1823	-8	-8	-8	-8	-8	8 -8	503	2	1	1	- 6	8 400	4			
4	4	First Russo	1	64	40 Ottoman I	- 2	2 4	26	1828	9	14	1829	-8	-8	-8	-8	-8	8 -8	506	11	2	2	14	8 80000	4			

2. Difficulties Faced (4)

4. Preserving UTF-8 Characters and File Size Handling

Many instances contained non-english characters (ex: Ferdinandsbrücke Project)

Microsoft Excel uses machine specific ANSI encoding

→ Had to use Google Spreadsheets to view UTF-8 characters

Google Spreadsheets does not support large amounts of data

- → just the Artwork csv files sum up to 25 MB
- → Artist table sums up to around 6 MB
- → raw MoMA csv file sums up to around 31 MB

3. Stored Procedures (1 MySQL Code)

1) Find the average size of 2D artworks during a certain war/conflict:

a)

Code

```
DELIMITER //
DROP PROCEDURE IF EXISTS AverageSizeDuringWar //
CREATE PROCEDURE AverageSizeDuringWar(IN title VARCHAR(20))
REGIN
  DECLARE foundWarName VARCHAR(20);
  DECLARE our CURSOR FOR
  SELECT WarName
  FROM Wars
  WHERE WarName = title;
  DECLARE EXIT HANDLER FOR NOT FOUND
    SELECT 'SORRY, not found' AS 'Error Message';
  OPEN cur;
  FETCH cur INTO foundWarName:
  CLOSE cur:
  SELECT A.year, ROUND(AVG(A.width * A.height). 2) as Average
  FROM Artwork A. Wars W
  WHERE W. WarName = foundWarName
    AND A.depth IS NULL
    AND A.year >= W.StartYear
    AND A.year <= W.EndYear
  GROUP BY A. YEAR:
END
```

3. Stored Procedures (1 Sample Run)

- 1) Find the average size of 2D artworks during a certain war/conflict:
 - b) Sample run

```
MySQL [cs41515_jchoi100_db]> call AverageSizeDuringWar('World War I');
+----+
| year | Average |
+----+
| 1914 | 3331.49 |
| 1915 | 840.89 |
| 1916 | 872.93 |
| 1917 | 1586.80 |
| 1918 | 1408.20 |
+----+
5 rows in set (0.29 sec)
```

3. Stored Procedures (2 MySQL Code)

2) Find Average size of 2D artworks created by artists from a certain nation:

a) Code

```
DROP PROCEDURE IF EXISTS AverageSizeFromNation //
CREATE PROCEDURE AverageSizeFromNation(IN nation VARCHAR(30), OUT size DECIMAL(10.2))
BEGIN
  DECLARE foundDemonym VARCHAR(30);
  DECLARE curl CURSOR FOR
  SELECT Demonvm
  FROM Demonvm
  Where nation = Country:
  DECLARE EXIT HANDLER FOR NOT FOUND
    SELECT 'SORRY, not found' AS 'Error Message';
 OPEN curl:
 FETCH curl INTO foundDemonvm:
 CLOSE cur1:
  SET size = (
  SELECT ROUND(AVG(A.width * A.height), 2) as Average
  FROM Artwork A, Artists At, WorkedOn W
  WHERE W.ArtistName = At.ArtistName
    AND W.ArtistBornYear = At.ArtistBornYear
    AND At.ArtistNationality = foundDemonym
    AND W.ObjectId = A.ObjectId
 GROUP BY At. ArtistNationality):
END
```

3. Stored Procedures (3 MySQL Code)

3) Compare the average size of 2D artworks from two nations.

Outputs are two average sizes and the ratio

a) Code

```
DROP PROCEDURE IF EXISTS CompareAverageSizeByNation //
CREATE PROCEDURE CompareAverageSizeByNation(IN nation1 VARCHAR(30), IN nation2 VARCHAR(30), OUT ratio DECIMAL(10, 2))
BEGIN
DECLARE foundSize1 DECIMAL(10, 2);
DECLARE foundSize2 DECIMAL(10, 2);

DECLARE EXIT HANDLER FOR NOT FOUND
SELECT 'SORRY, not found' AS 'Error Message';

CALL AverageSizeFromNation(nation1, @size1);
CALL AverageSizeFromNation(nation2, @size2);

SET foundSize1 = (SELECT @size1);
SET foundSize2 = (SELECT @size2);

SET ratio = ROUND((foundSize1 / foundSize2), 2);
END
//
```

3. Stored Procedures (3 Sample Run)

3) Compare the average size of 2D artworks from two nations.

Outputs are two average sizes and the ratio

b) Example Run

3. Stored Procedures (4 MySQL Code)

- 4) Find The number of Artworks/Films from a certain nation purchased or donated :
 - a) Code

```
DROP PROCEDURE IF EXISTS ShowCreditByArtistNation //
CREATE PROCEDURE ShowCreditByArtistNation(IN nation VARCHAR(30), OUT numPurchase INTEGER, OUT numGift INTEGER)
BEGIN
DECLARE foundDemonym VARCHAR(30);

DECLARE curl CURSOR FOR
SELECT Demonym
FROM Demonym
Where nation = Country;

DECLARE EXIT HANDLER FOR NOT FOUND
SELECT 'SORRY, not found' AS 'Error Message';

OPEN curl;
FETCH curl INTO foundDemonym;
CLOSE curl;
```

3. Stored Procedures (4 MySQL Code cont'd)

(Code Continued)

```
SET numPurchase = (
   SELECT SUM(K.count) as sum
    FROM ((
     SELECT count(A.ObjectId) as count
     FROM Artwork A. Artists At. WorkedOn W
     WHERE A.ObjectId = W.ObjectId
        AND W.ArtistName = At.ArtistName
        AND W.ArtistBornYear = At.ArtistBornYear
        AND At.ArtistNationality = foundDemonvm
        AND A.CreditLine LIKE '%Purchase%')
    INTON (
     SELECT count(F.ObjectId) as count
     FROM Film F. Artists At. WorkedOn W.
      WHERE E.ObjectId = W.ObjectId
        AND W.ArtistName = At.ArtistName
        AND W.ArtistBornYear = At.ArtistBornYear
        AND At.ArtistNationality = foundDemonym
        AND F.CreditLine LIKE '%Purchase%')) as K);
  SET numGift = (
   SELECT SUM(K.count) as sum
    FROM ((
     SELECT count(A.ObjectId) as count
     FROM Artwork A, Artists At, WorkedOn W
     WHERE A.ObjectId = W.ObjectId
        AND W.ArtistName = At.ArtistName
        AND W.ArtistBornYear = At.ArtistBornYear
        AND At.ArtistNationality = foundDemonvm
        AND A. CreditLine LIKE '%Gift%')
    UNION (
      SELECT count(F.ObjectId) as count
     FROM Film F. Artists At. WorkedOn W
      WHERE F.ObjectId = W.ObjectId
        AND W.ArtistName = At.ArtistName
        AND W.ArtistBornYear = At.ArtistBornYear
        AND At.ArtistNationality = foundDemonvm
        AND F. CreditLine LIKE '%Gift%')) as K);
END //
```

4. Advanced Topics / What We Have Learned

Advanced Topics

- Parsing real world data
- Dealing with large amounts of data
- Extracting interesting information/correlation from various domains

4. Advanced Topics / What We Have Learned

What Did We Learn?

- Parsing is difficult!
- Parsing real world data takes time and a lot of error handling
- There are a lot of interesting facts/trends/conjectures one can extract from large amounts of data

5. Future Development

- Natural Language Processing
 - a. Given a query that involves aggregation, be able to parse table information in more detail
 - ex) User input: Show me whose name appears most frequently as a donor/contributor to MoMA's collection

Some tuple columns:

- Max Underwood Fund
- Gift of Max Underwood
- Purchase and partial gift of the architect in honor of Max Underwood
- → Then, recognize that all of these tuple columns point to a single donor:

Max Underwood

Thank you!