# 600.315 Demo Presentation

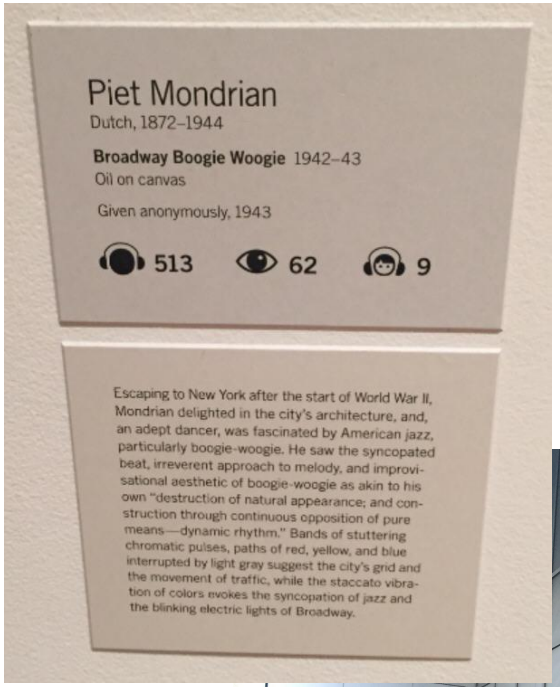
Joon Hyuck Choi (jchoi100),
Joo Chang Lee (jlee381)

# Contents

# 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





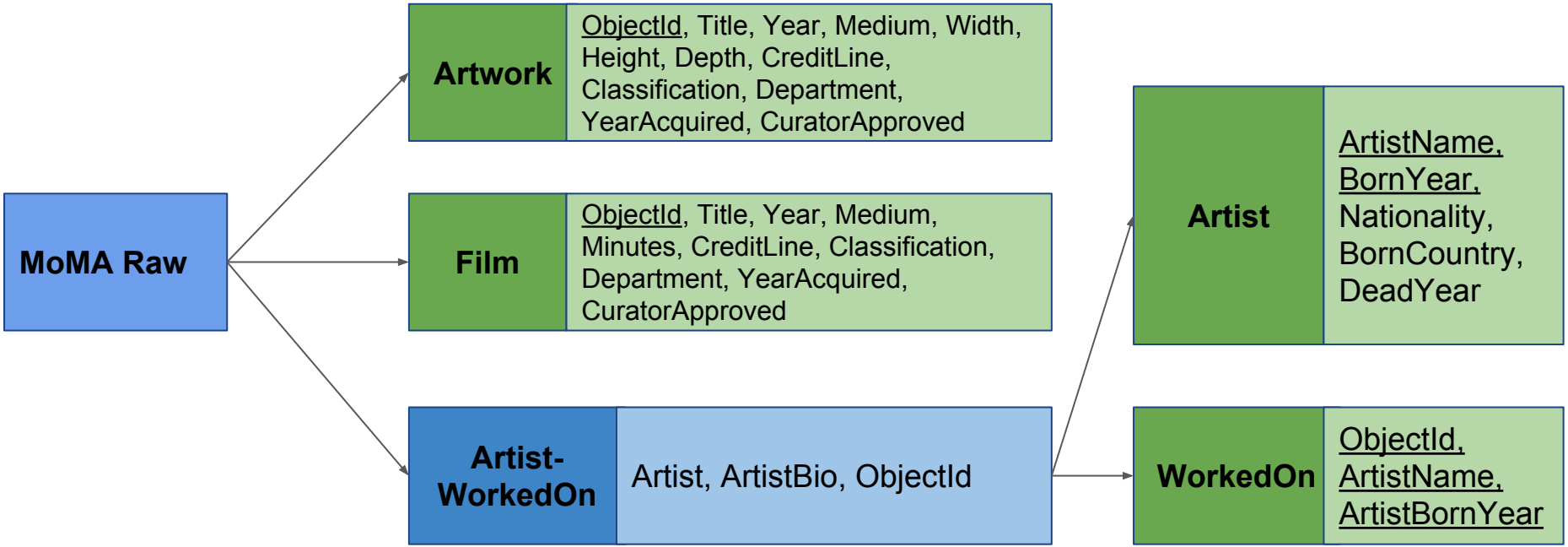
# 1. Overview / Domain (continued)

1. MoMA Database (Raw artwork.csv)

| Title | Artist | ArtistBio | Date | Medium | Dimensions | CreditLine | MoMANumber | Classification | Department | DateAcquired | Curator Approved | ObjectID | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 472.1941 | Painting | Painting & Sculpture | 1941-06-10 | Y | 79802 | http: |
| The Red Studio | Henri Matisse | (French, 1869-1954) | Issy-les-Moulineaux, fall 1911 | 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, Paris | Rem Koolhaas, Madelon Vriesendorp | (Dutch, born 1944) (Dutch, born 1945) | 1987 | Synthetic polymer pa | 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: |

# 1. Overview / Domain (continued)

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



# 1. Overview / Domain (continued)

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

Extra-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | ccode1 | SideA | ccode2 | SideB | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | Initiator | Interven | TransFror | Outcome | TransTo | WhereFou | BatDeath | NonState | Version |
| 2 | 300 | Allied Bon | 3 | 210 | Netherlan | -8 | -8 | 8 | 26 | 1816 | 8 | 30 | 1816 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 1 | -8 | 1 | -8 | 6 | 13 | -8 | 4 |
| 3 | 300 | Allied Bon | 3 | 200 | United Kir | -8 | Algeria | 8 | 26 | 1816 | 8 | 30 | 1816 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 1 | -8 | 1 | -8 | 6 | 129 | 6000 | 4 |
| 4 | 301 | Ottoman- | 3 | 640 | Ottoman I | -8 | Saudi Wal | 9 | -9 | 1816 | 9 | 11 | 1818 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 0 | -8 | 1 | -8 | 6 | 13500 | 14000 | 4 |

Intra-StateWarData_v4.1.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | CcodeA | SideA | CcodeB | SideB | Intnl | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | TransFror | WhereFou | Initiator | Outcome | TransTo | SideADea | SideBDea | Version |
| 2 | 500 | First Cauc | 5 | 365 | Russia | -8 | Georgians | 0 | 6 | 10 | 1818 | -9 | -9 | 1822 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 2 | Chechens | 1 | -8 | 5000 | 6000 | 4.1 |
| 3 | 501 | Sidon-Dar | 6 | -8 | Sidon | -8 | Damascus | 0 | 6 | -9 | 1820 | 7 | 21 | 1821 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 6 | Sidon | 2 | -8 | -9 | -9 | 4.1 |
| 4 | 502 | First Two | 4 | 300 | Austria | -8 | -8 | 1 | 3 | -9 | 1821 | 3 | 23 | 1821 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 2 | Liberals | 1 | -8 | -9 | -8 | 4.1 |

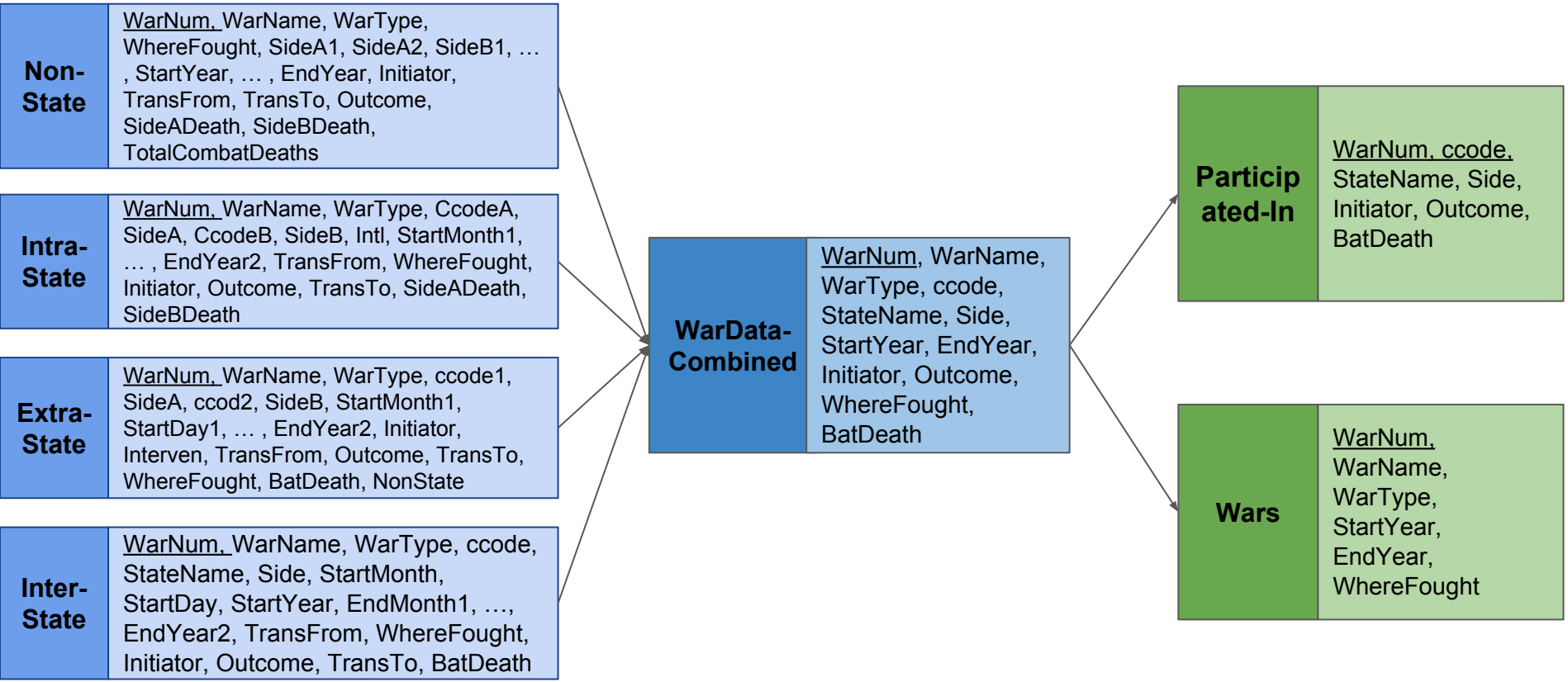
Non-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | WhereFou | SideA1 | SideA2 | SideB1 | SideB2 | SideB3 | SideB4 | SideB5 | StartYear | StartMon | StartDay | EndYear | EndMonth | EndDay | Initiator | TransFror | TransTo | Outcome | SideADea | SideBDea | TotalCom | Version | | | |
| 2 | 1500 | First Maor | 8 | 9 | Te Raupar | -8 | Taranaki | Ngai Tahu | Waikato | Ngati Ira | Rangitikei | 1818 | -9 | -9 | 1824 | -9 | -9 | A | -8 | -8 | 1 | 1500 | 6000 | 7500 | 4 | | | |
| 3 | 1501 | Shaka Zul | 8 | 4 | Shaka Zul | -8 | Bantu | -8 | -8 | -8 | -8 | 1819 | -9 | -9 | 1828 | 9 | 24 | A | -8 | -8 | 1 | 20000 | 40000 | 60000 | 4 | | | |

Inter-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | ccode | StateNan | Side | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | TransFror | WhereFou | Initiator | Outcome | TransTo | BatDeath | Version | | | |
| 2 | 1 | Franco-Sp | 1 | 230 | Spain | 2 | 4 | 7 | 1823 | 11 | 13 | 1823 | -8 | -8 | -8 | -8 | -8 | -8 | 503 | 2 | 2 | 2 | -8 | 600 | 4 | | | |
| 3 | 1 | Franco-Sp | 1 | 220 | France | 1 | 4 | 7 | 1823 | 11 | 13 | 1823 | -8 | -8 | -8 | -8 | -8 | -8 | 503 | 2 | 1 | 1 | -8 | 400 | 4 | | | |
| 4 | 4 | First Russ | 1 | 640 | Ottoman I | 2 | 4 | 26 | 1828 | 9 | 14 | 1829 | -8 | -8 | -8 | -8 | -8 | -8 | 506 | 11 | 2 | 2 | -8 | 80000 | 4 | | | |

# 1. Overview / Domain (continued)



# 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
        i. 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
        i. 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), l. 16' 2" x w. 31 1/2" (l. 492.8 x w. 80 cm)

# 2. Difficulties Faced (3)

## 3. Combining (unioning) tables with incompatible domain types

a. ex)

Extra-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | ccode1 | SideA | ccode2 | SideB | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | Initiator | Interven | TransFror | Outcome | TransTo | WhereFou | BatDeath | NonState | Version |
| 2 | 300 | Allied Bon | 3 | 210 | Netherlan | -8 | -8 | 8 | 26 | 1816 | 8 | 30 | 1816 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 1 | -8 | 1 | -8 | 6 | 13 | -8 | 4 |
| 3 | 300 | Allied Bon | 3 | 200 | United Kir | -8 | Algeria | 8 | 26 | 1816 | 8 | 30 | 1816 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 1 | -8 | 1 | -8 | 6 | 129 | 6000 | 4 |
| 4 | 301 | Ottoman- | 3 | 640 | Ottoman I | -8 | Saudi Wal | 9 | -9 | 1816 | 9 | 11 | 1818 | -8 | -8 | -8 | -8 | -8 | -8 | 1 | 0 | -8 | 1 | -8 | 6 | 13500 | 14000 | 4 |

Intra-StateWarData_v4.1.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | CcodeA | SideA | CcodeB | SideB | Intnl | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | TransFror | WhereFou | Initiator | Outcome | TransTo | SideADea | SideBDea | Version |
| 2 | 500 | First Cauc | 5 | 365 | Russia | -8 | Georgians | 0 | 6 | 10 | 1818 | -9 | -9 | 1822 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 2 | Chechens | 1 | -8 | 5000 | 6000 | 4.1 |
| 3 | 501 | Sidon-Dar | 6 | -8 | Sidon | -8 | Damascus | 0 | 6 | -9 | 1820 | 7 | 21 | 1821 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 6 | Sidon | 2 | -8 | -9 | -9 | 4.1 |
| 4 | 502 | First Two | 4 | 300 | Austria | -8 | -8 | 1 | 3 | -9 | 1821 | 3 | 23 | 1821 | -8 | -8 | -8 | -8 | -8 | -8 | -8 | 2 | Liberals | 1 | -8 | -9 | -8 | 4.1 |

Non-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | WhereFou | SideA1 | SideA2 | SideB1 | SideB2 | SideB3 | SideB4 | SideB5 | StartYear | StartMon | StartDay | EndYear | EndMonth | EndDay | Initiator | TransFror | TransTo | Outcome | SideADea | SideBDea | TotalCom | Version | | | |
| 2 | 1500 | First Maor | 8 | 9 | Te Raupar | -8 | Taranaki | Ngai Tahu | Waikato | Ngati Ira | Rangitikei | 1818 | -9 | -9 | 1824 | -9 | -9 | A | -8 | -8 | 1 | 1500 | 6000 | 7500 | 4 | | | |
| 3 | 1501 | Shaka Zul | 8 | 4 | Shaka Zul | -8 | Bantu | -8 | -8 | -8 | -8 | 1819 | -9 | -9 | 1828 | 9 | 24 | A | -8 | -8 | 1 | 20000 | 40000 | 60000 | 4 | | | |

Inter-StateWarData_v4.0.csv

| | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | WarNum | WarNam | WarType | ccode | StateNan | Side | StartMon | StartDay | StartYear | EndMonth | EndDay1 | EndYear1 | StartMon | StartDay | StartYear | EndMonth | EndDay2 | EndYear2 | TransFror | WhereFou | Initiator | Outcome | TransTo | BatDeath | Version | | | |
| 2 | 1 | Franco-Sp | 1 | 230 | Spain | 2 | 4 | 7 | 1823 | 11 | 13 | 1823 | -8 | -8 | -8 | -8 | -8 | -8 | 503 | 2 | 2 | 2 | -8 | 600 | 4 | | | |
| 3 | 1 | Franco-Sp | 1 | 220 | France | 1 | 4 | 7 | 1823 | 11 | 13 | 1823 | -8 | -8 | -8 | -8 | -8 | -8 | 503 | 2 | 1 | 1 | -8 | 400 | 4 | | | |
| 4 | 4 | First Russ | 1 | 640 | Ottoman I | 2 | 4 | 26 | 1828 | 9 | 14 | 1829 | -8 | -8 | -8 | -8 | -8 | -8 | 506 | 11 | 2 | 2 | -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))
BEGIN
  DECLARE foundWarName VARCHAR(20);

  DECLARE cur 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 cur1 CURSOR FOR
  SELECT Demonym
  FROM Demonym
  Where nation = Country;

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

  OPEN cur1;
  FETCH cur1 INTO foundDemonym;
  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**

```
MySQL [cs41515_jchoi100_db]> call CompareAverageSizeByNation('France', 'England'
, @ratio);
Query OK, 0 rows affected (0.48 sec)

MySQL [cs41515_jchoi100_db]> select @size1 as size1, @size2 as size2, @ratio as
ratio;
+---------+--------+-------+
| size1   | size2  | ratio |
+---------+--------+-------+
| 1186.83 | 135.28 |  8.77 |
+---------+--------+-------+
1 row in set (0.00 sec)
```

# 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 cur1 CURSOR FOR
  SELECT Demonym
  FROM Demonym
  Where nation = Country;

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

  OPEN cur1;
  FETCH cur1 INTO foundDemonym;
  CLOSE cur1;
```

# 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 = foundDemonym
        AND A.CreditLine LIKE '%Purchase%')
    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 = 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 = foundDemonym
        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 = foundDemonym
        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
     i. 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!