A Name of Thrones

INF 385T: Data Wrangling, SQL & Beyond
Dr. James Howison

Nathan M. Heep
nmh872
December 4 2016

Files (all csvs in same dir):
~/project/prototype/projectimport.py
~/project/prototype/prototype.py

This project analyses relationships between U.S. baby names that occur more than 5 times per year since 1981 according to U.S. census data, and character names from the popular fantasy book series A Song of Ice and Fire (ASOIAF) and HBO entertainment series Game of Thrones (GOT). The goal of this project is to determine if ASOIAF or GOT have influenced baby naming trends since their releases in 1996 and 2011, respectively.

The analysis considers a time scale of 34 years from 1981 to 2015, beginning 15 years prior to ASOIAF's release (1981), to the release of ASOIAF (1996), to the release of GOT (2011), and finally to the final year of the US baby name dataset, 2015. The name dataset contains the year, name, and number of people born with that name that year. Name data were compared against a dataset of all character names from the ASOIAF/GOT universe. This file was analysed for popular names from specific houses that serve as the central cast of characters. The results from this analysis suggest that ASOIAF/GOT has influenced U.S. baby name trends since their respective releases in 1996 and 2011.


We Do Not Sow
Greyjoy

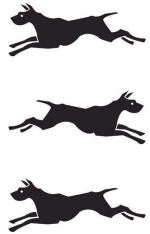

Fire and Blood
Targaryen


Winter is Coming
Stark


Hear Me Roar
Lannister


Ours is the Fury
Baratheon


Unbowed, Unbent, Unbroken
Martell


Brothers Clegane & Other
Minor characters

PROCESS

The following steps were necessary to import and transform data for analysis. All code and comments are located in projectimport.py and prototype.py.

- Import all US baby names to holden (iSchool server) after the year 1980. This provides a sense of naming trends for the fifteen years prior to ASOIAF's release. Original file had year data extending only to 2014.
- Locate and prep the 2015 baby name file by adding a header row, year column, and appending 2015 to all rows.
- Import all character names to holden from the ASOIAF/GOT universe.
- Strip all first names from the GOT table and append to a new field "fixedname" with only first names.
- Output names to a new table "gotnames" on holden.
- Compare gotnames with national_names table, append new column called inGOT, insert 0 , 1 if the name is not inGOT or is inGOT.
- Output CSV that contains all US baby names from all years that match a GOT name for that year.
- Import CSV to Tableau 10.1 for visualization and analysis.

Balon, Theon, Asha

Daenerys, Nymeria, Balor, Aemon, Brynden Rivers

Arya, Robb, Catelyn, Sansa, Bran, Jon Snow

Tywin, Tyrion, Jaime

Renly

Oberyn

Brothers Sandor & Gregor Clegane with Minor Characters Beric, Edric, Ashara, Shae, Jorah

WORKFLOW SCREENSHOTS

Database / Tables

characters

id	name	title	house	heir	spouse	fixedname
1	Viserys II Targaryen			Aegon IV Targaryen		Viserys
2	Walder Frey	Lord of the Crossing	House Frey		Perra Royce	Walder
3	Addison Hill	Ser	House Swyft			Addison
4	Aemma Arryn	Queen	House Arryn		Viserys I Targaryen	Aemma
5	Sylva Santagar	Greenstone	House Santagar		Eldon Estermont	Sylva
6	Tommen Baratheon			Myrcella Baratheon		Tommen
7	Valarr Targaryen	Hand of the King	House Targaryen		Kiera of Tyrosh	Valarr
8	Viserys I Targaryen			Rhaenyra Targaryen		Viserys
9	Wilbert	Ser				Wilbert
10	Wilbert Osgrey	Ser	House Osgrey			Wilbert
11	Will		Night's Watch			Will

Has a first name from a book

* ... 1

Derived from full name

gotnames

fixedname
Viserys
Walder
Addison
Aemma
Sylva
Tommen
Valarr
Wilbert
Will
Willam
Willem
Willifer
Willit
Willis
Willow
Willum
Wolmer

Was given a first name at birth

1 ... *

Matches a character name

national_names

id	name	year	gender	count
1	Jennifer	1981	F	57041
2	Jessica	1981	F	42526
3	Amanda	1981	F	34372
4	Sarah	1981	F	28166
5	Melissa	1981	F	28006
6	Amy	1981	F	20339
7	Nicole	1981	F	20312
8	Stephanie	1981	F	20207
9	Elizabeth	1981	F	20171
10	Heather	1981	F	17943
11	Michelle	1981	F	17770
12	Rebecca	1981	F	16733
13	Kimberly	1981	F	16691
14	Tiffany	1981	F	16416

SQL/PYTHON EXPORT

Query to create a new table that joins gotnames and national_names where the two match, along with the number of people with that name for that year.

```
SELECT national_names.name AS Name, national_names.year AS Year, count AS 'Number of
People', IF( gotnames.fixedname IS NULL , 0, 1 ) AS inGOT
FROM national_names
LEFT OUTER JOIN gotnames ON national_names.name = gotnames.fixedname
HAVING inGOT = 1
```

Name	Year	Number of People	inGOT
Melissa	1981	28006	1
Holly	1981	4565	1
Robin	1981	2868	1
Bethany	1981	2693	1
Barbara	1981	2531	1
Jaime	1981	2386	1
Beth	1981	2310	1
Joanna	1981	1825	1
Joy	1981	1621	1
Marissa	1981	1450	1
Robyn	1981	1367	1
Rhonda	1981	1356	1
Alyssa	1981	1129	1
Johanna	1981	880	1
Jocelyn	1981	809	1
Emma	1981	533	1

Python code to run the query and output to a new file, nationalgotonlyoutput.csv

```
with connection.cursor() as cursor:
  sql = """
SELECT national_names.name AS Name, national_names.year AS Year, count AS 'Number of
People', IF( gotnames.fixedname IS NULL , 0, 1 ) AS inGOT
FROM national_names
LEFT OUTER JOIN gotnames ON national_names.name = gotnames.fixedname
HAVING inGOT = 1
  """
  cursor.execute(sql)
  results = cursor.fetchall()
  csv_column_order = list(results[0].keys())

  with open('nationalgotonlyoutput.csv', 'w', newline='') as csvfile:
    myCsvWriter = csv.DictWriter(csvfile, delimiter=',',
                                 # quotechar='"',
                                 # fieldnames = csv_column_order)
    myCsvWriter.writeheader()
    for row in results:
      myCsvWriter.writerow(row)
```

Year	Number o	Name	inGOT
1981	28006	Melissa	1
1981	4565	Holly	1
1981	2868	Robin	1
1981	2693	Bethany	1
1981	2531	Barbara	1
1981	2386	Jaime	1
1981	2310	Beth	1
1981	1825	Joanna	1
1981	1621	Joy	1

Data was imported into Tableau for analysis

Analysis
This section provides a series of graphs representing how characters names from prominent families that match U.S. baby names trend according to U.S. census data. The gray area represents all years after ASOIAF's 1996 release, while the dotted line represents the release of HBO's GOT television show.

House Stark

Results suggest that ASOIAF/GOT has influenced baby name trends in two instances, **Arya** and **Sansa**.

Sansa did not appear in U.S. census rolls from 1981 to 2011, and was first introduced in 2012 with 11 names, increasing marginally each year to 20 names in 2015.

Arya was a name that appeared in U.S. census data from 1981 to 1996 with around 20 persons with that name per year. The name grew in popularity since 96' rising to 498 in GOT's release year. Since that point, the name has substantially increased in U.S. census rolls, nearly doubling the following year eventually to count of 1,732 in 2015.

Jon, however, suffers the inverse. The name regresses from 2,077 instances in 1981 to 2015 instances in 328. Clearly, Jon Snow has not done enough for his namesake.

House Lannister

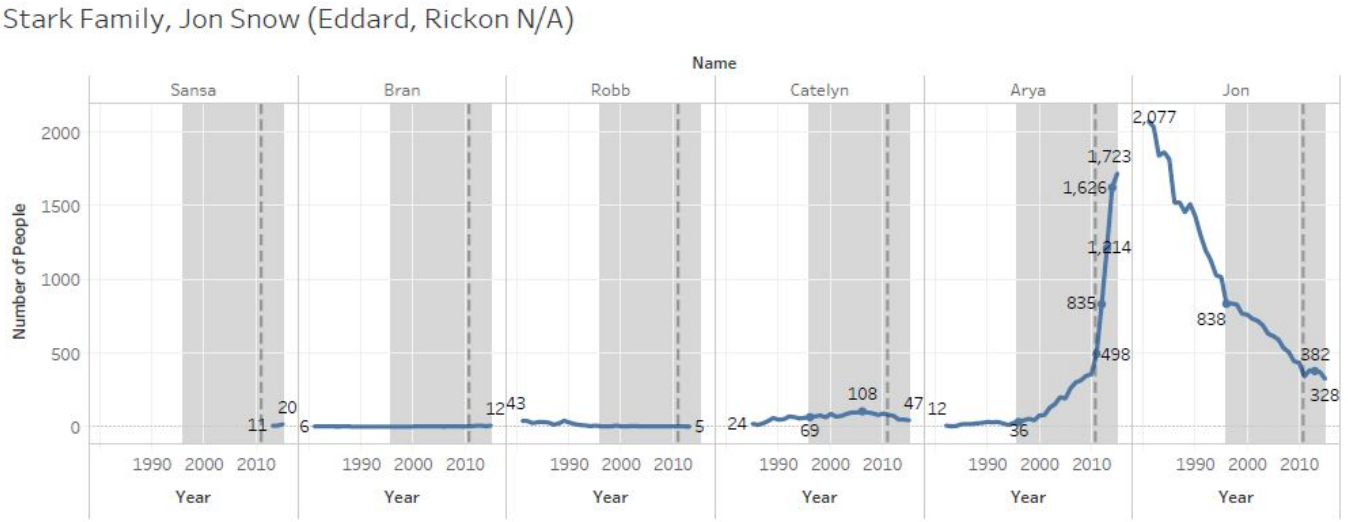
Results suggest that the popularity of ASOIAF/GOT has influenced baby name trends in two instances, **Tywin** and **Tyrion.**

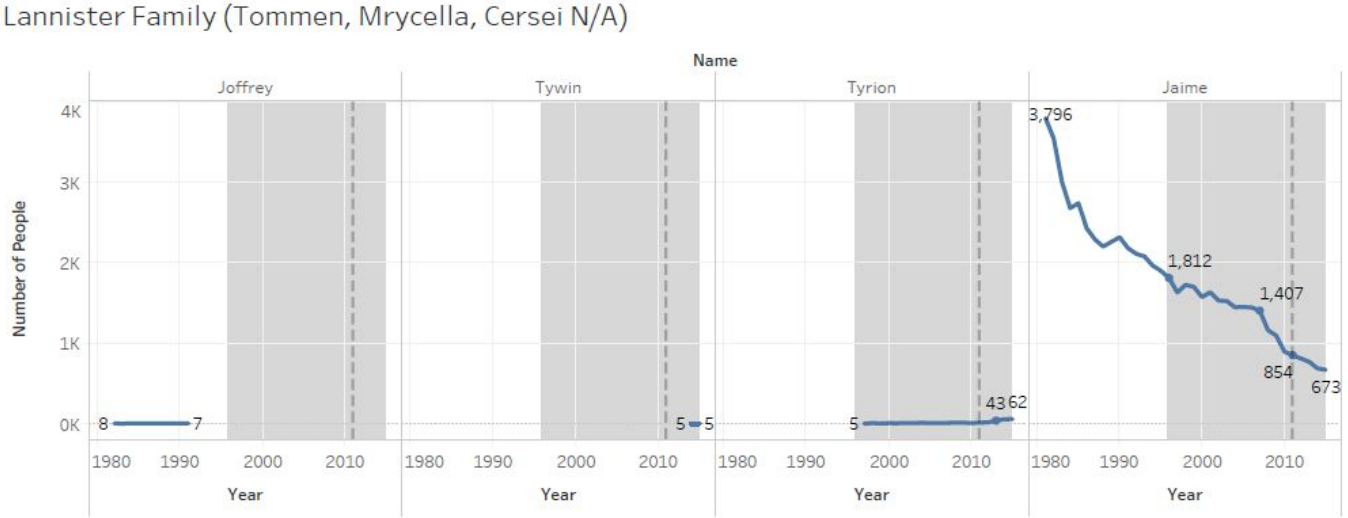
Joffrey was in use from 1981 to 1991, but fell out of favor 5 years before ASOIAF's release, never to return.

Tywin first entered U.S. rolls in 2014 and continued in 2015, only introduced after that character appears in the GOT show.

Tyrion first entered the census in 1997, one year after ASOIAF's release. This figure increased from 2011 to 2015 after GOT's release.

Similar to Jon, **Jaime** suffers a regressing trend. Once again the character fails the name.





House Greyjoy

Results are not conclusive if ASOIAF/GOT have influenced baby naming trends surrounding the Greyjoy family.

Balon was last used in 1981 and not again after that year.

Theon existed as a name prior to ASOIAF/GOT release, but gradually increased from 2011 to 2015. This may indicate the show's popularity has had an influence.

Asha was more popular than her kin prior to 1981 but results are not clear if ASOIAF/GOT has had any influence. In the show, this character's name changed to Yara and was not present in any of the U.S. census data.

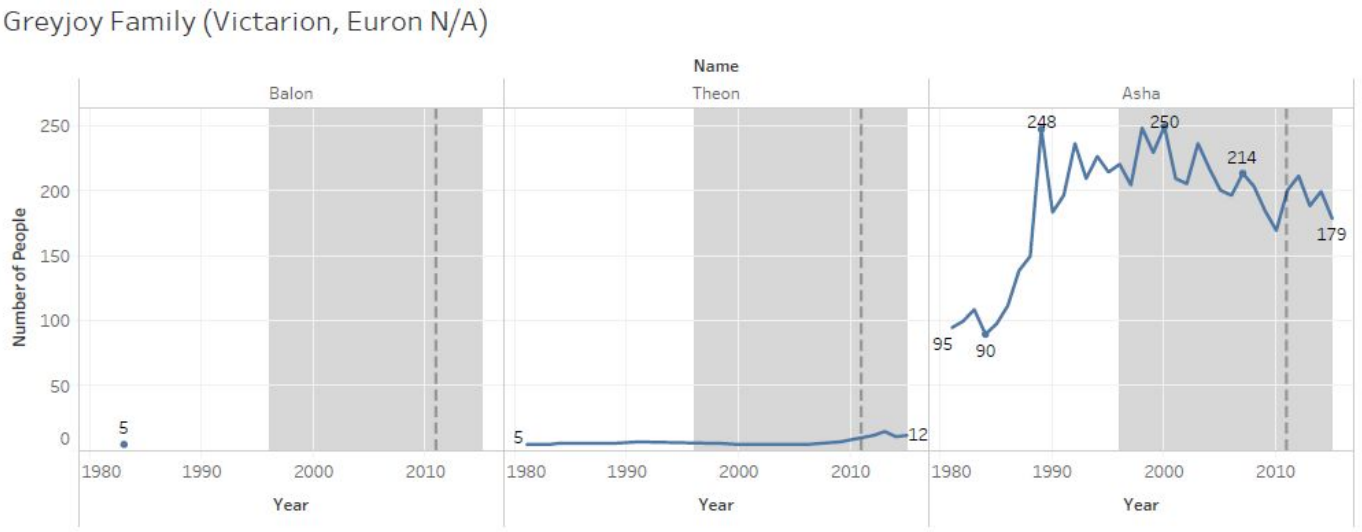
House Targaryen

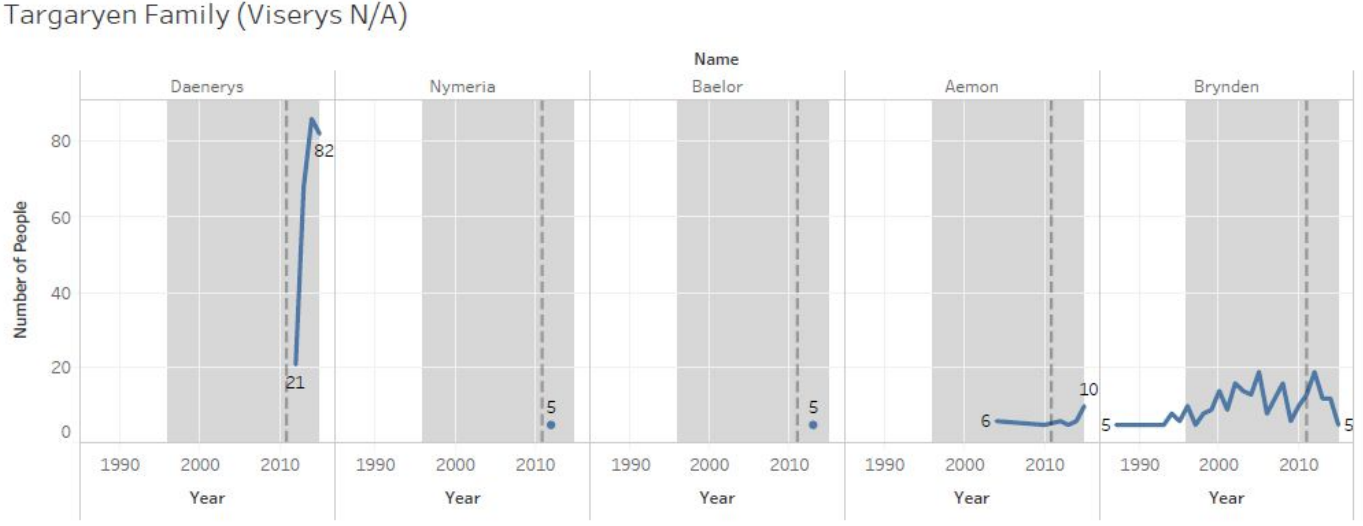
Results suggest that the popularity of Targaryen names have potentially influenced baby name trends in four instances, **Daenerys**, **Nymeria**, **Baelor**, and **Aemon**.

Daenerys is one of the most popular characters in the ASOIAF universe, and although her name is introduced with ASOIAF's release, the name does not appear in the census rolls until 2011 with GOT's debut. The name has increased by around 20 names per year since 2011, ending with 82 instances in 2015.

Nymeria and **Baelor** are both introduced after the GOT show's release, and although are minor Targaryen characters in the show, their names are used as pet names, location names, and characters names throughout the ASOIAF/GOT universe.

Brynden Rivers is a bastard with Targaryen blood, and real name of the Three-Eyed Raven. Although his name was used prior to ASOIAF/GOT's release, it experiences periods of increase and decrease after the 1996. It is not conclusive if ASOIAF/GOT have influenced trends around this name.





Brothers Clegane, Shae, Oberyn

Results are not conclusive if ASOIAF/GOT have influenced baby naming trends surrounding **Sandor** Clegane, **Gregor** Clegane, or **Shae**. Results do indicate that **Oberyn** has potentially influenced naming trends.

Sandor and **Gregor** were both names uses prior to the 1996 release, but show now signs of significant increase or decrease after ASOIAF/GOT's release.

Oberyn, an extremely popular character after his appearance in GOT, first appears in the US census rolls in 2015 with 14 instances.

House Targaryen

Results suggest that the ASOIAF/GOT universe has influenced naming trends in three instances, **Beric**, **Jorah**, and **Renly**.

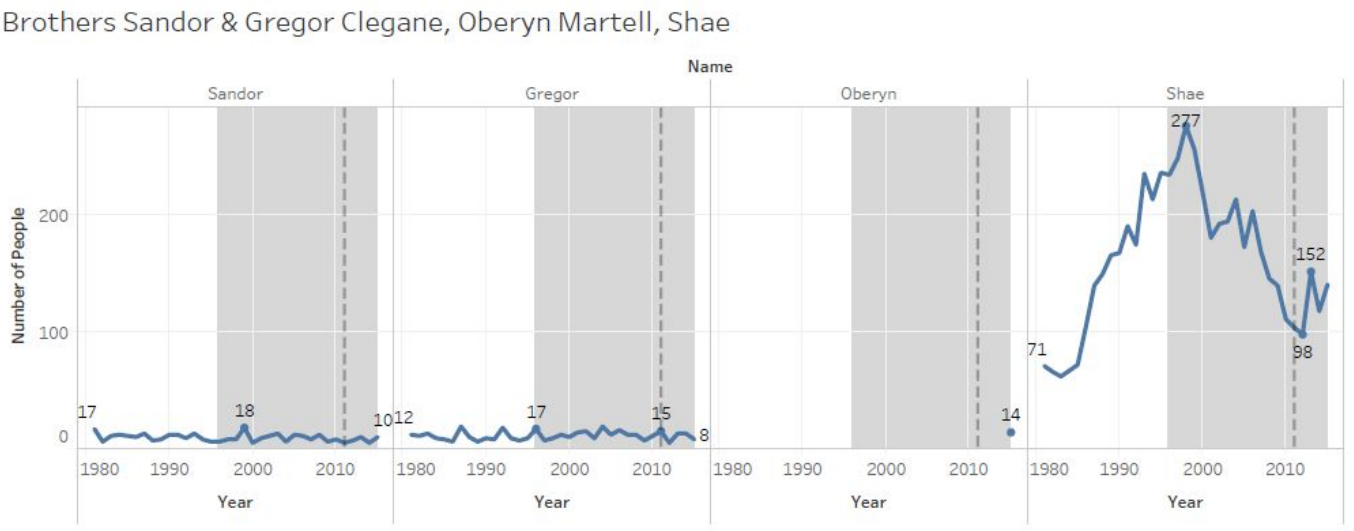
Ashara was used as a name prior to 1996, but does not show variance after ASOIAF/GOT release that signals any influence.

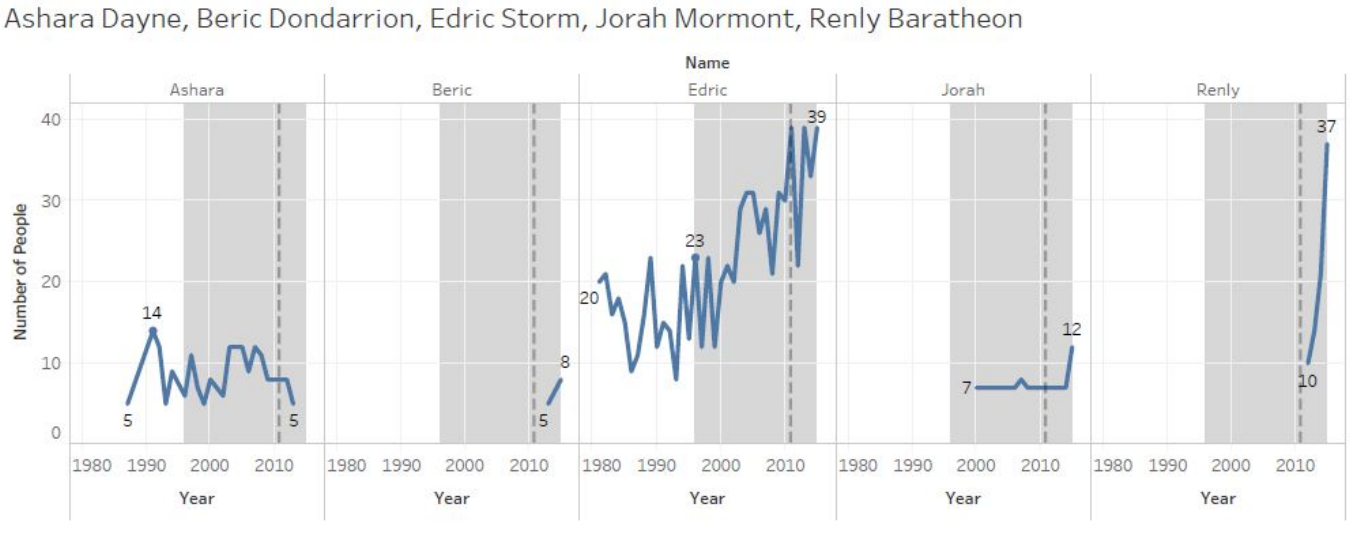
Beric did not enter U.S. census rolls until 2012, a year after that character's introduction in GOT.

Similar to Ashara, **Edric** appears as a name prior to ASOIAF/GOT and although there is significant variance, there is not enough information to determine the characters have influenced naming trends.

Jorah appears as a name first in 2000, four years after ASOIAF's release. This name increases in popularity from 2013 to 2015.

Renly does not enter the census until 2012, one year after GOT's release with 10 instances, increasing by around 10 each year to a total of 37 in 2015.





CONCLUSION

The results from this analysis suggest that character names from the ASOIAF universe have influenced naming trends for the following names: **Arya**, **Sansa**, **Daenerys**, **Nymeria**, **Tywin**, **Tyrion**, **Renly**, **Oberyn**, and **Jorah**.

CHALLENGES

This was a very fun project that presented a number of challenges.

- Merging all 50 different state name files into one text file, then importing that data. It ended up being very easy but took a while to figure out. Ironically, this never found its way to the final analysis.
- Appending the header, a new column, and populating the data with 2015 was important to get the national names table up to date.
- Stripping the GOT first name from the last name allowed me to join the two tables together. Prior to this I was unsure how I would get the two tables to join.

In general, I really enjoyed learning:

- How to transform data on import via python/sql
- How to get tables talking together in python/mysql and export the results to a CSV

VISUALIZATION

I first tried using **R** and **ggplots2** to try to visualize my data, but was never able to get **R** to function the way I wanted. I intend to explore **R** in the future to visualize this data, but for the purposes of the assignment I transitioned to using **Tableau**.

Using **Tableau**, I imported the csv that matched inGOT names with national names for my analysis. I created a variety of graph and chart types that presented the data in interesting ways, using reference lines and bands to identify attributes of the graph. The tool is very intuitive and allows you to manipulate a visualization in a similar manner as a pivot table.

EPILOGUE

Thanks for reading.