

# MAT259 Proj 3: Data Correlation

Song Gao

Geography Department (Email: [sgao@geog.ucsb.edu](mailto:sgao@geog.ucsb.edu))

Website: <http://www.geog.ucsb.edu/~sgao/>

## Concept

The Presidential Election is one of the most important political events in United States. The presidential debates present the party's visions during the campaign. The major issues debated were the economy and jobs, the federal budget deficit, taxation and spending, the future of social security, medicare, healthcare reform, education, social issues, immigration, and foreign policy. The debates are very valuable during the public vote process. It is interesting to explore the temporal trends of check-outs behaviors related to "election" or "debate" in Seattle Public Library (SPL) and the articles about these topics in New York Times (NYT) during the election periods, as well as to identify whether there is a correlation pattern between the two public media.

## Visual Design and Color

I will use the comparison-mode visual design for the correlation project which means every single visual element should contain the binary information between two media (SPL v.s. NYT), such as the bar charts, the frequency display and the relative weight. For the color design, I am going to use the Germany Flag color scheme of **Red**, **Yellow** and **Black**.

## Data and Query

### (1) Monthly check-outs from "2008-01-01" to the date "2012-12-31"

```
SELECT year(cout), month(cout), count(*)
```

```
FROM inraw
```

```
WHERE (title like "%election%" or "%debate%") and date(cout)>"2008-01-01" AND  
date(cout)<"2012-12-31"
```

```
GROUP BY year(cout),month(cout)
```

```
ORDER BY year(cout),month(cout);
```

**Processing Time:** 57.409 seconds and 60 rows returned

## Results

[http://www.geog.ucsb.edu/~sgao/mat259/2008\\_2012Monthly\\_electiondebate\\_checkout\\_s.csv](http://www.geog.ucsb.edu/~sgao/mat259/2008_2012Monthly_electiondebate_checkout_s.csv)

**(2) Daily check-outs from “2008-01-01” to the date “2012-12-31”**

```
SELECT year(cout),dayofyear(cout), count(*)
```

```
FROM inraw
```

```
WHERE (title like "%election%" or "%debate%") and date(cout)>"2008-01-01" AND  
date(cout)<"2012-12-30"
```

```
GROUP BY year(cout),dayofyear(cout)
```

```
Order BY year(cout),dayofyear(cout);
```

**Processing Time:** 57.236 seconds and 1738 rows returned

## Results

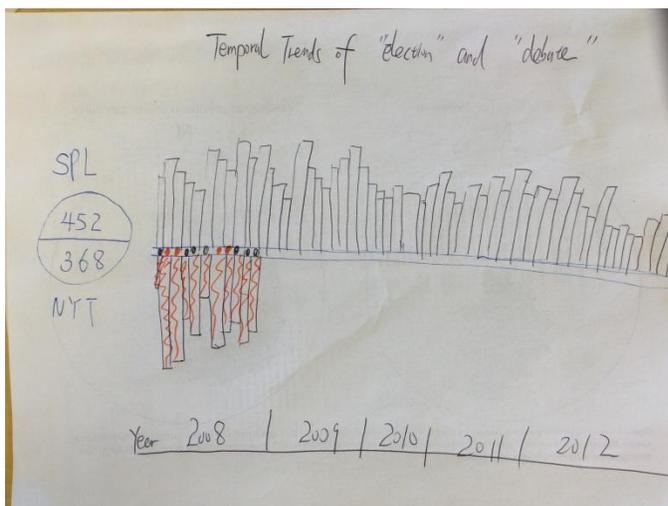
[http://www.geog.ucsb.edu/~sgao/mat259/2008\\_2012Daily\\_electiondebate\\_checkouts.csv](http://www.geog.ucsb.edu/~sgao/mat259/2008_2012Daily_electiondebate_checkouts.csv)

And the daily check-outs data specifically for the Election Year 2008 and 2012:

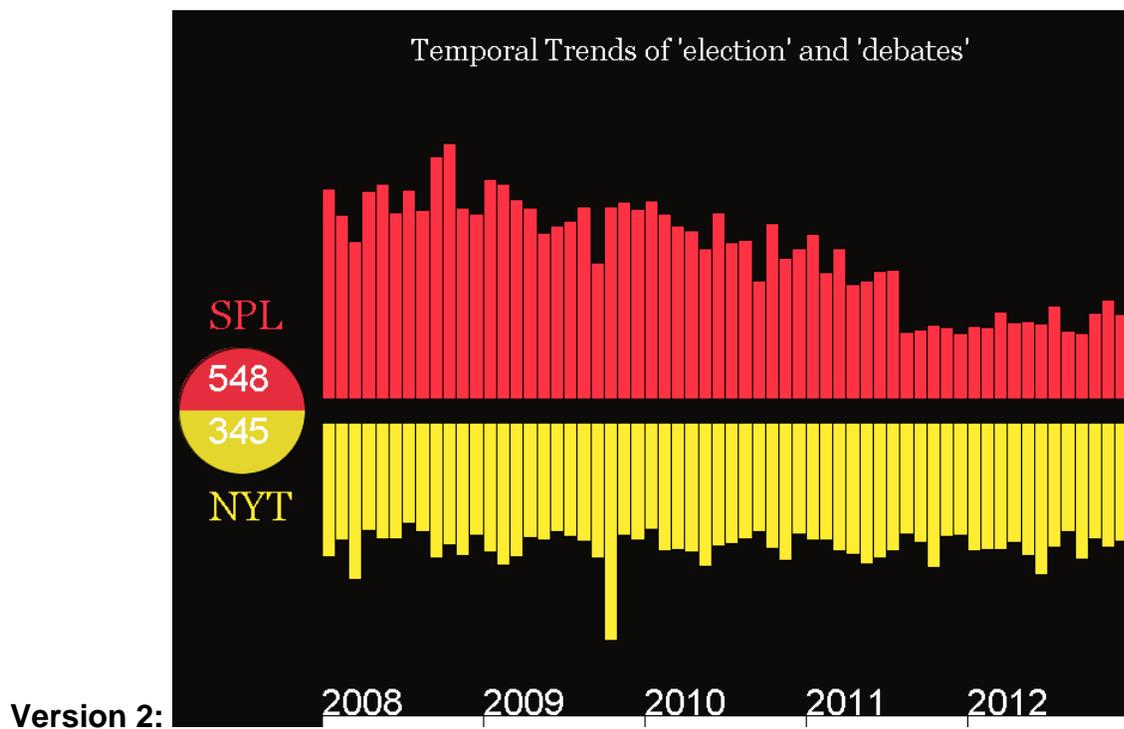
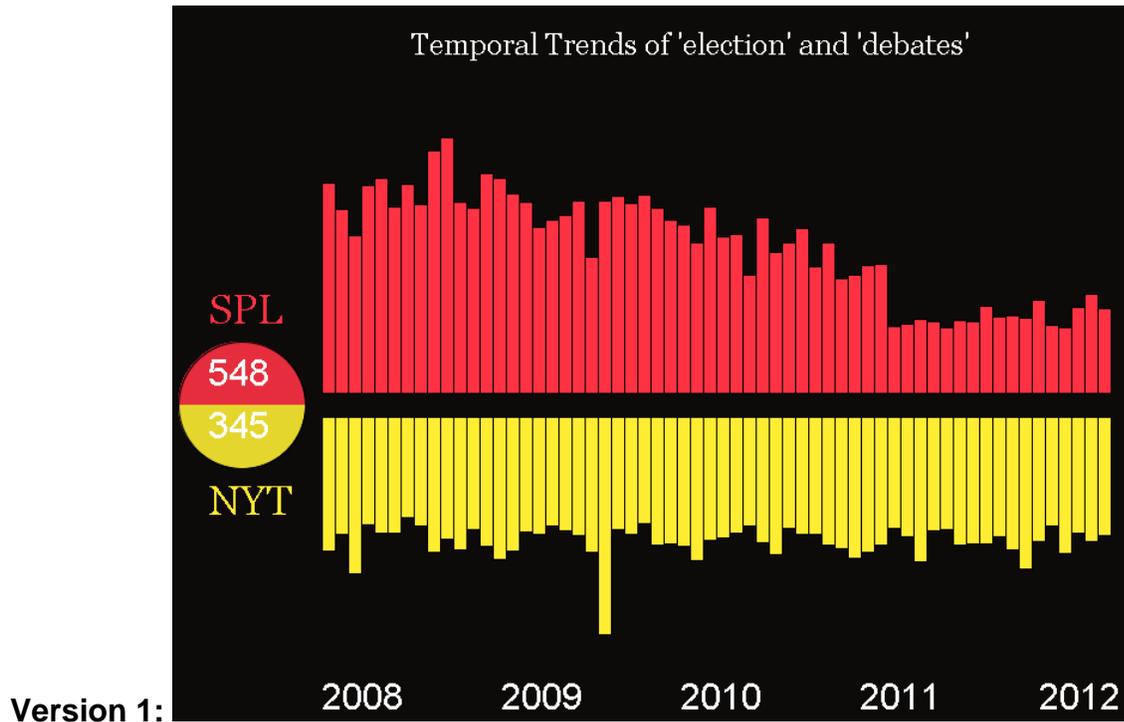
[http://www.geog.ucsb.edu/~sgao/mat259/2008Daily\\_electiondebate\\_checkouts.csv](http://www.geog.ucsb.edu/~sgao/mat259/2008Daily_electiondebate_checkouts.csv)

[http://www.geog.ucsb.edu/~sgao/mat259/2012Daily\\_electiondebate\\_checkouts.csv](http://www.geog.ucsb.edu/~sgao/mat259/2012Daily_electiondebate_checkouts.csv)

Doodle:



## Results



It also supports the interaction by hovering the mouse on the bars and the half circles will show the changing values of monthly check-outs in Seattle Public Library and articles in New York Times containing the keywords: “%*election*%” or “%*debate*%”.

Version 3: (with real New York Times Data loaded in)

