

# Data & Big Data

## Assignment 3 II Correlating data

### By Mohit Hingorani

#### Concept:

Correlating data to me means making sense of two data sets at the same time, and understanding the dynamics between the two sets. For this assignment I am exploring the trends in the emerging fields of "Data & Big Data". I will be using Article Search from the New York Times API and looking for the words 'Data' and a subset of it 'Big Data'. In the Seattle Public Library, I will be searching for books with the titles containing the words 'Data' & 'Big Data'. Very interesting trends and patterns emerge.

I will be looking into data from (2011-2013): a 3-year period.

#### Labels:

I am going for a Swiss poster design style: featuring bold fonts and a minimalistic design. The data is arranged vertically (like a time line) instead of the usual horizontal flow. I will be using a combination of simple bars & lines for the project. The mouse position will highlight the number of books/ articles that the bar represents for both SPL & NYT.

#### Color Scheme:

Red & White

#### Query:

#### SQL queries:

```
select month(cout),year(cout),sum(case when itemtype = "acbk" or "jcbk" then 1 else 0 end) as book
from inraw where title like "% data %" and date(cout) >= "2011-01-01" and date(cout) <= "2014-03-03"
group by month(cout),year(cout) order by year(cout), month(cout)
```

```
select month(cout),year(cout),sum(case when itemtype = "acbk" or "jcbk" then 1 else 0 end) as book
from inraw where title like "%big data %" and date(cout) >= "2011-01-01" and date(cout) <= "2014-03-03"
group by month(cout),year(cout) order by year(cout), month(cout)
```

#### New York Times queries:

#### Example request

String request = "";

```
request += http://api.nytimes.com/svc/search/v2/articlesearch + ".json";
```

```
request += "?q=" + big+data;
```

```
request += "&facet_field=source";
```

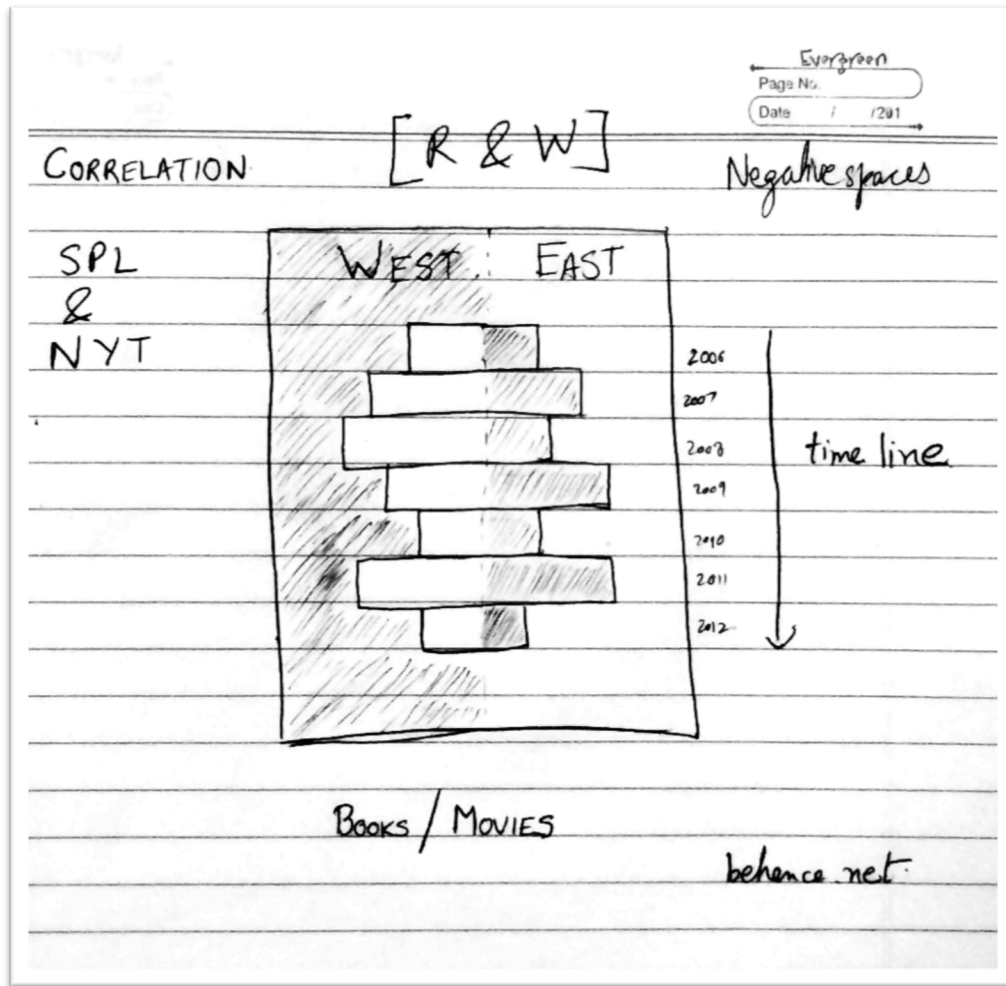
```
request += "&facet_filter=true";
```

```
request += "&begin_date=" + 20140101;
```

```
request += "&end_date=" + 20140131;
```

request += "&api-key=" + 362412324bfd4d15c3937caa04125f28:9:68817159;

Doodle:



### Results:

The results were pretty striking. One can see an exponential increase in the NYT article search for 'Data' in August of 2012. The results on the 'Big Data' double during the same month.

A corresponding increase is noted for the Seattle Public Library as well. There is a noticeable increase for 'Data' related books in August 2013: an 8-month time difference. What is interesting is that the first checkout for the first Big Data related book happens in September 2013: One full year after the field of data (and by extension big data) exploded in the computer science industry. It brings in interesting questions for the publishing industry highlighting the time taken to write a technical book on a new emerging field and the trends and readership of books and articles on emerging technology fields.

# DATA & BIG DATA

Data Correlation:

SPL: Checkouts in books with the word 'Data' & 'Big Data'

NYT: Number of times 'data' & 'big data' turned up in the article search

2011

2011

2012

2012

2013

2013

Seattle Public Library

New York Times

By Mohit Hingorani || MAT259 || Assignment 3

