

Report for DS4D Assignment 3

-- The visibility of place in Dundee's newspaper

S1620916 Yan WU

S1631436 Yifan CHEN

S1770406 Chiao-Jung Lo

1. Context

Every newspaper article trying to report an event should cover "Five W's": Who, What, Why, Where, When and sometimes How and So what (Geoff-hart.com, 2017).

The place data is very necessary in the articles. Some places become visible because the media report the things happened there. As a result, people start to know that place. For example, the Chinese missing student news in Edinburgh was also reported by lots of Chinese media. Chinese people all posted or reposted the news article through the social media. Now in China, many people knew and remembered the place, Edinburgh, through this news (see figure 1).



Figure 1

Like famous people are payed attention by public media, place also has its own popularity, and the popularity can change with time. Then we generate our idea: try to find the popularity of places in 1950's Dundee. At that time, people had limited access to gain the information and maybe the newspaper was one of the most popular method to get information around them. From the newspaper's articles, Dundee's people can know the big events and where it happened. Therefore, the place frequency on the newspaper, to some degree, can represent the visibility of that place. Our group would like to explore more about the visibility of places in Dundee's newspaper and try to find some relationship between newspaper, events and places.

2. Data

2.1 Data analysis

Our group got the place data in 1950's Dundee newspaper article from our data holder. Our original data (see figure 2) has places' names, location, dates and articles' titles. From assignment 2, each of us analyzed the same data from the different direction and tried to find the relationship between the place frequency and titles, the place frequency and places' names, the place frequency and dates. Then, we found some interesting things and had some hypothesizes from our data.

```
In [2]: import pandas as pd
import re
data = pd.read_table("Dundee1950-January.tsv")
print(data.head())
```

| | place | lat | long | title | date |
|---|----------------|---------|---------|-------------------------------|------------|
| 0 | DUNDEE | 56.4584 | -2.9737 | SOCIALIST WINDOW-DRESSING | 1950-01-03 |
| 1 | Harriet Street | 56.4717 | -2.9589 | Bawbee Bets Now A Penny | 1950-01-03 |
| 2 | Dundee | 56.4584 | -2.9737 | WAS 50 YEARS IN JUTE INDUSTRY | 1950-01-03 |
| 3 | Gateside | 56.5870 | -2.9148 | WAS 50 YEARS IN JUTE INDUSTRY | 1950-01-03 |
| 4 | Dudhope Street | 56.4649 | -2.9741 | WAS 50 YEARS IN JUTE INDUSTRY | 1950-01-03 |

Figure 2

2.2 Hypothesis

2.2.1 Large cities

We thought relatively large cities were often under the spotlight. We compute the top 20 frequency of city names in news title as our analyzed target. Those cities (see figure 3) are more massive scale and population. Some are the historical harbor, international commercial areas, and important cities.

| date | Arboath | Forfar | Montrose | Brechin | Carnoustie | Monifieth | BroughtyFerry | Kirriemuir | Tannadice | Glamis | Craigie | Edzell | Tayport | EastEnd | StCyrus | In |
|------------|---------|--------|----------|---------|------------|-----------|---------------|------------|-----------|--------|---------|--------|---------|---------|---------|----|
| 1950-08-01 | 5 | 5.0 | 6.0 | 2.0 | 1.0 | NaN | 1.0 | 4.0 | NaN | 1.0 | NaN | 2.0 | 1.0 | NaN | NaN | |
| 1950-08-03 | 2 | 2.0 | 4.0 | 6.0 | 1.0 | NaN | 2.0 | NaN | NaN | 2.0 | NaN | NaN | 1.0 | NaN | 2.0 | |
| 1950-08-04 | 5 | 4.0 | 3.0 | 4.0 | 3.0 | 4.0 | 1.0 | 1.0 | NaN | NaN | 1.0 | NaN | NaN | NaN | NaN | |
| 1950-08-05 | 4 | NaN | 5.0 | 1.0 | NaN | 2.0 | NaN | 1.0 | 1.0 | NaN | NaN | 2.0 | NaN | NaN | NaN | |
| 1950-08-07 | 8 | 7.0 | 4.0 | 2.0 | 5.0 | 5.0 | 2.0 | NaN | 1.0 | NaN | 3.0 | 2.0 | 1.0 | 1.0 | 1.0 | |
| 1950-08-08 | 1 | 5.0 | 4.0 | 3.0 | 3.0 | 2.0 | 1.0 | NaN | 1.0 | NaN | 1.0 | NaN | NaN | NaN | NaN | |

Figure 3

2.2.2 Sports events

We thought sports events might be very popular in Dundee, 1950. Through counting the number of frequent words of news title for the most common top 200 words, some sports-related words appear many times, such as "League"(33 times), "Rugby"(33 times), "Golf"(26 times) and so forth.

From bar charts comparing total number of daily news and that of sports news. We found every two or three days, at least a piece of sports news was reported

and places in Dundee, or local people who would like to learn more about historical news report of their hometown. The idea is that every place has its own popularity in media (newspaper here), and visitors can see how much the media payed attention on that place in past time.

The basic point we try to convey is the change of popularity for each chosen place with time. We especially pay attention on the peak point: Which day is the most frequent? What happened on earth on that day? Also, we try to explain the reason of each dramatically changed point. Here, we use the title of article to explain what happened on earth for that specific day instead of the content of the whole article. Though we did find out the data through British Newspaper Archive, showing the copy of article seems like a redundant method. Our basic idea is to keep it simple but meaningful. For these titles, some of them could be big events like "Mearns Church's 102nd Anniversary" which could be reported by many newspapers. Others seem not so interesting and could be reported just by one newspaper, but they did happen on that day. Anyway, these two factors affect the dramatic change points or the peak points.

4. Demo

The prototype combines both author-driven and reader-driven approaches, i.e. Martini Glass structure (see figure 5) to tell our story (Segel and Heer, 2010). It begins with a data comic where some questions, observations are used to introduce the idea of the visualization. Once the author's intended narrative is complete, the reader-driven part engages user to interactively explore the data.

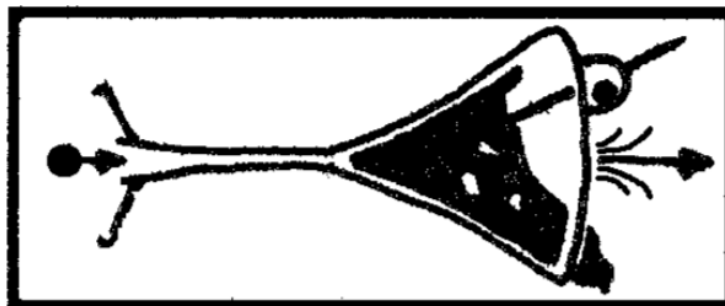


Figure 5

For data comic part (see figure 6), we drew the overview of our physical artefact and explained all parts of this artefact. Then we told audience how to interact with this artefact.

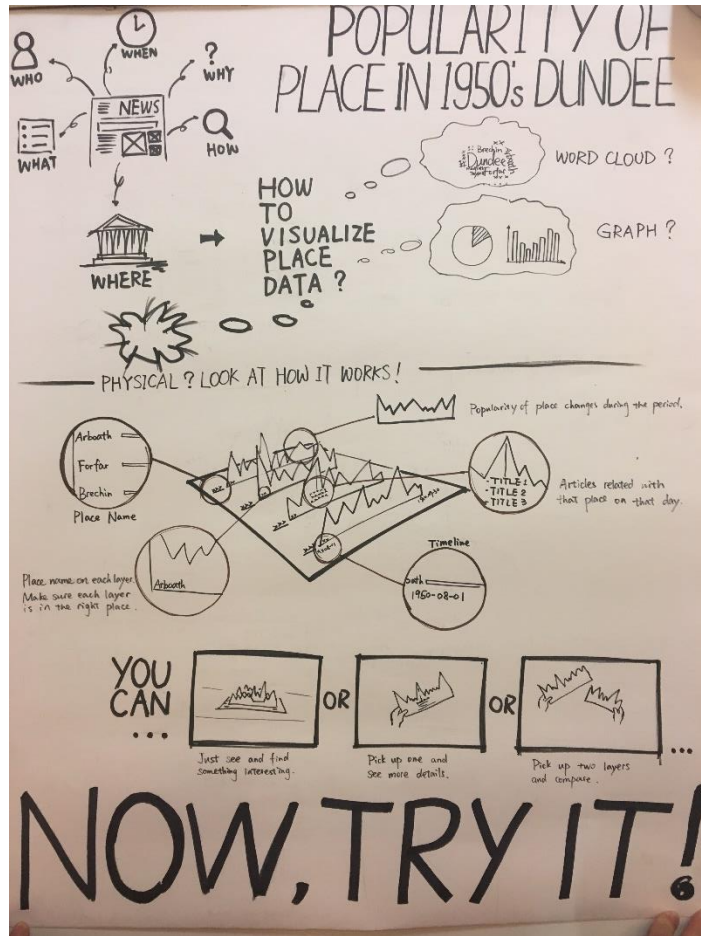


Figure 6

For the physical artefact (see figure 7 and figure 8), we used laser-cutting machine to cut each layer on the MDF boards and also, engraved the titles on some layers. We also created a wooden base to install all these layers, but all layers are flexible and audience can move each layer.

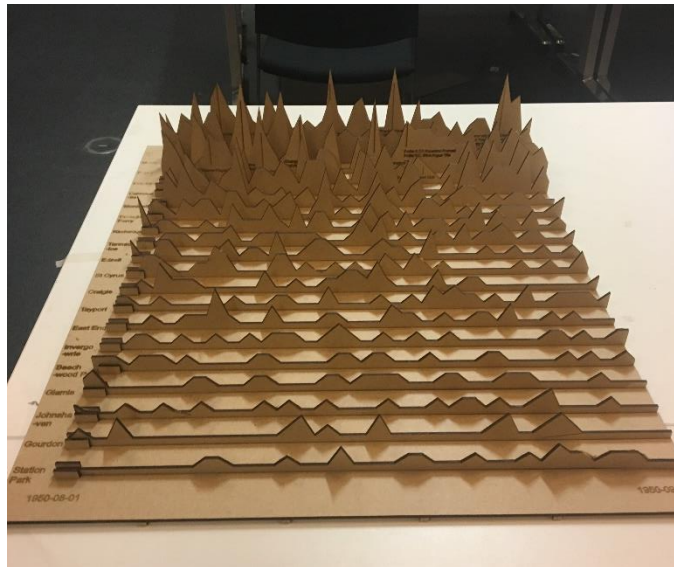


Figure 7



Figure 8

Reference

Geoff-hart.com. (2017). The five w's of online help systems. [online] Available at: <http://www.geoff-hart.com/articles/2002/fivew.htm> [Accessed 4 Dec. 2017].

En.wikipedia.org. (2017). Brechin. [online] Available at: <https://en.wikipedia.org/wiki/Brechin> [Accessed 4 Dec. 2017].

Segel, E. and Heer, J. (2010). Narrative Visualization: Telling Stories with Data. IEEE Transactions on Visualization and Computer Graphics, 16(6), pp.1139-1148.