# Visualization of Settlement Change in the Chifeng Region (Northeastern China)

Qiao Zhang, Junhui Chen

Computer Science Department
University of Pittsburgh

Tao Li

Department of Anthropology
University of Pittsburgh

**Abstract**

To help the study of human activities from earliest times in China, far outside the Central Plain, our team aims to visualize the density data estimated from pottery types found in the local area, classified and organized in a chronological order, which can provide archaeologists an intuitive way of observation.

**Keywords**: visualization, settlement change, Chifeng, northeastern China

## 1 Introduction

Chifeng international collaborative archaeological research project (subsequently, the Chifeng Project) is a project conducted by a group of archaeologists from China and foreign countries. They documented patterns of change in settlement in the Chifeng region by conducting a systematic regional field survey that aims to begin to explain those changes. Based on that, they are able to analyze the change of human density in certain areas.

## 2 Theory

Density of a large area is represented by density of the sites within. Estimation the density of larger area is not that straightforward, and frankly speaking, not accurate enough, so they divide that area into small regions, density of each region is estimated and thus the density of a certain site surrounded by regions can also be estimated.

The measurement of human density, especially from prehistory time, can never be an easy task. Archaeologists manage to collect pottery fragments from different sites and place them in chronological order (which called seriation), so the distribution of a certain time can be modeled, thus the density of human at that time, since those fragments are the proof of people once living there, and they can be seen as correlated.

### 2.1 Original Implementation

Originally, domain experts use static visualization to learn occupation for each time period. As illustrated in Figure 1 and Figure 2, they use static pictures to demonstrate each time period's density. This implementation is limited because the whole Chifeng area is large and the details are lost on overview.

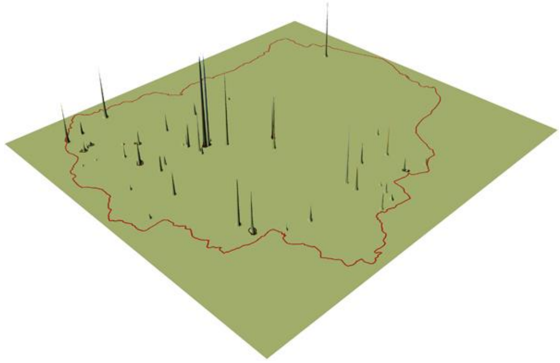


Figure 1: Original static visualization: unsmoothed density surface for Xiaoheyan occupation.

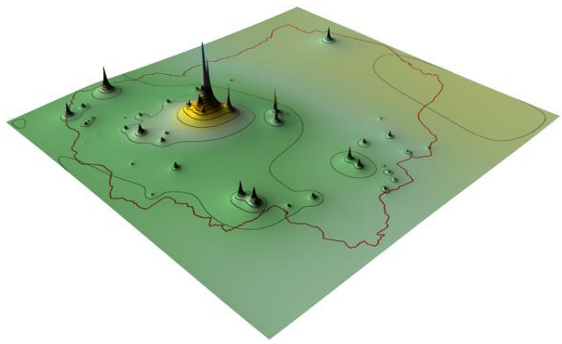
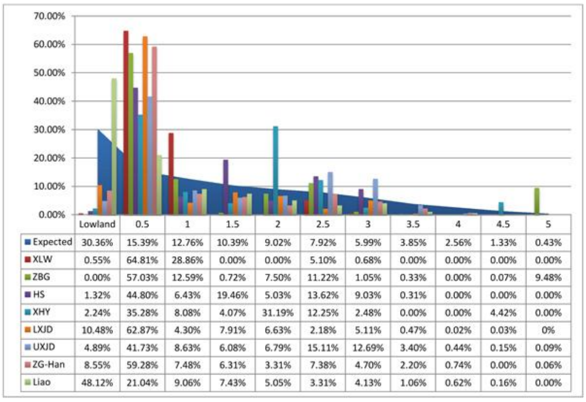


Figure 2: Original static visualization: smoothed density surface for Xiaoheyan occupation.

Domain experts also use bar charts to visualize the data. However, as we all know, it is not straightforward and somehow cluttered, which is hard to interpret.



| | Lowland | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Expected | 30.36% | 15.39% | 12.76% | 10.39% | 9.02% | 7.92% | 5.99% | 3.85% | 2.56% | 1.33% | 0.43% |
| XLW | 0.55% | 64.81% | 28.86% | 0.00% | 0.00% | 5.10% | 0.68% | 0.00% | 0.00% | 0.00% | 0.00% |
| ZBG | 0.00% | 57.03% | 12.59% | 0.72% | 7.50% | 11.22% | 1.05% | 0.33% | 0.00% | 0.07% | 9.48% |
| HS | 1.32% | 44.80% | 6.43% | 19.46% | 5.03% | 13.62% | 9.03% | 0.31% | 0.00% | 0.00% | 0.00% |
| XHY | 2.24% | 35.28% | 8.08% | 4.07% | 31.19% | 12.25% | 2.48% | 0.00% | 0.00% | 4.42% | 0.00% |
| LXJD | 10.48% | 62.87% | 4.30% | 7.91% | 6.63% | 2.18% | 5.11% | 0.47% | 0.02% | 0.03% | 0% |
| UXJD | 4.89% | 41.73% | 8.63% | 6.08% | 6.79% | 15.11% | 12.69% | 3.40% | 0.44% | 0.15% | 0.09% |
| ZG-Han | 8.55% | 59.28% | 7.48% | 6.31% | 3.31% | 7.38% | 4.70% | 2.20% | 0.74% | 0.00% | 0.06% |
| Liao | 48.12% | 21.04% | 9.06% | 7.43% | 5.05% | 3.31% | 4.13% | 1.06% | 0.62% | 0.16% | 0.00% |

Figure 3: Original bar chart used for visualization.

## 2.2 Objectives

The original visual implementations are not fully utilizing data visualization as a powerful tool. First, the original pictures are static, with no animation or interaction. Second, no systematic tool is implemented, thus domain experts cannot intuitively and effectively browse through data. Third, the visualizations of statistics information is disconnected with the geographical information.

A series of snapshots cannot be good enough to represent an invisible dimension which is time. Or even more, like comparison between different time periods, digging in for detailed information. A proper visualization method is needed for these kinds of requirements.

Our goal is to build an effective and intuitive visualization tool that allows the users, i.e. domain experts, to explore the data in an interactive way. We utilized several visualization principles, such as layering, overview and details, details on demand, etc.

## 2.3 Visualization

### 2.3.1 Methodology decision

The prototype is made after a heated discussion. We first browsed the D3 gallery, looking for inspiration. At first, we thought we could use streamgraph to visualize our fragments data because data variation along with the time is what the users are most interested in. Streamgraphs are perfect for visualizing time variations. However, because stream graph needs interpolation between each time period, we could not assume that between each time period there would be any pottery unearthed, hence, in order not to mislead the users, we gave up stream graphs.

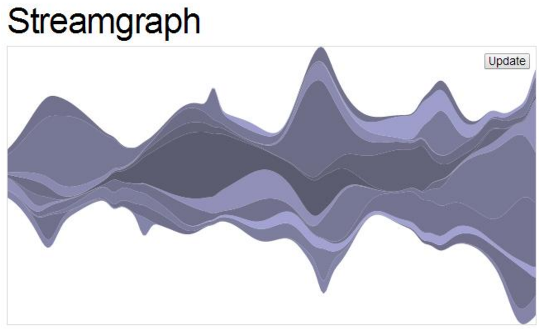



Figure 4: Streamgraph.

We then considered using treemap, because it is easy to represent weight information (unearthed pottery amount). However, the geographical information is hard to represent using treemap and treemap is most proper to present hierarchical data, which is obviously not our case.

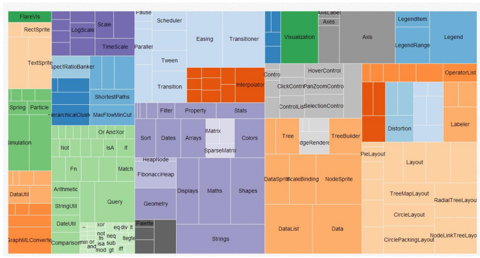


Figure 5: Treemap.

Then we decided to use basic scatter plot as our final decision. Because scatter plot has inherent strength in representing coordinate system, and node size can be representing weights. Also it is relatively easy to integrate the fourth dimension, time, into the visualization, using interaction and animation. Also, in order to meet the consultant's needs for details, we decided to use details on demand method to update a detail view when mouse hovers a circle.

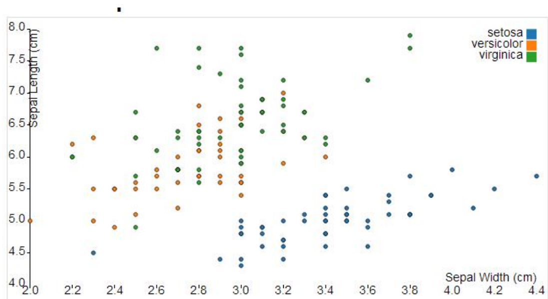



Figure 6: Scatter Plot.

The reason we used layering is because we want to give proper feedback to the users. Even though we divided the whole region into smaller ones, it is still possible that multiple circles lay on each other. Thus, we added a hover effect that dims the other circles out of focus to help the users distinguish the current circle.

Also, we implemented details on demand technique that allows interactively selecting parts of data to be visualized more detailed while providing an overview of the whole informational concept. After clicking a specific circle, the detail view will be updated and the information of unearthed pottery fragment amount along with time is displayed. Moreover, if the user hovers her mouse on the circles in detail view, a window will pop out, displaying the corresponding time period name and the exact number.

### 2.3.2 Data processing

As for the ceramics, the original data comes in tabular format, as below:

| | |
|---|---|
| 1 | Key--A combination of the survey tract number, the team letter, and the collection unit number for a unique text string that identifies each collection unit (corresponds to the text string that identifies the collection unit on the maps) |
| 2 | Site--The number of the site to which the collection unit belongs |
| ... | ... |

Table 1: Listing of Survey Collection Units(KSD)

| | |
|---|---|
| 1 | Key--a combination of the survey tract number, the team letter, and the collection unit number for a unique text string that identifies each collection unit |
| 2 | Total number of sherds1 |
| 3 | Total number of sherds2 |
| ... | ... |

Table 2: Ceramics from Survey(CT)

While for the geographical data, dxf format is used, which can be processed by AutoCAD. With EATTEXT command, the original data can be further exported into tabular format. The output format is shown in Table 3.

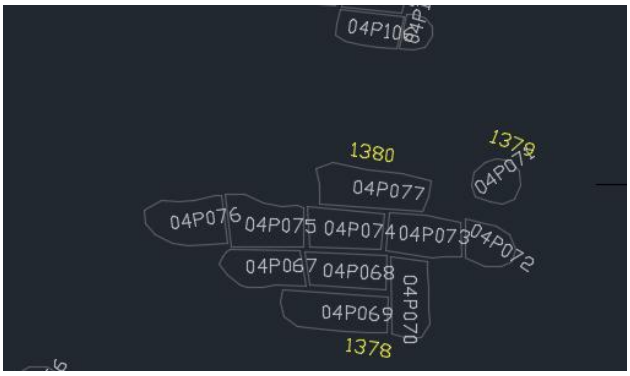



Figure 7: Geographical data

| | |
|---|---|
| 1 | Key or Site |
| 2 | Coordinate |
| ... | ... |

Table 3: Coordinates(COOD)

We use site as the smallest component in our visualization, which means site number should be our “key” instead of units in the original dataset. We could merge all the data into a huge structure, which would fit our requirements, but with future improvement taken into consideration, we decided to make least change to the original data structure, and dynamically extracted data from the 3-part dataset and generated temporal structure as needed, which later turned out to be a good idea when we notice that under the circumstance of mapping all sites on a single coordinate system, important details are inevitably lost, and too distracting for the user as well. Thus we decided to cut down the amount of data shown to the user at the same time, which is to divide the whole dataset into smaller parts, and show as demanded. But it doesn’t mean the overview is completely useless, since more or less it show some overall trend.

Since the data is generated run-time, an effective access method is needed to meet the requirement of frequent reads. As a result, we construct 3 hash indexes as below:

```
For each item in COOD
    If it is Key
        K2A[Key]=KSD[Key] U CT[Key]
        If Site Exists
            S2K[Site]+=Key
        Else
            S2K[Site]=[Key]
    Else
        If Site Exists
            S2K[Site]+=Key
        Else
            S2K[Site]=[Key]

        If GroupNumber Exists
            G2SK[GroupNumber]+=Key or Site
        Else
            G2SK[GroupNumber] = [Key] or [Site]
```

```
K2A(Key)={GroupNumber,SiteNumber,X,Y,sherds1,…,sherdN}
S2K(SiteNumber)={GroupNumber,Keys[],X,Y}
G2SK(GroupNumber)=SiteNumbers[]
```

With the help from these 3 hash indexes, the generation of data for certain group is straightforward, Find all the sites that belong to a group, then calculate the total count of each site based on those units that belong to it. Each item in the temporal structure will be like this:

```
{SiteNumber,GroupNumber,Counts[[Period,Count]],Xs[[Period,X]],Ys[[Period,Y]]}
```

## 3 User Study

We conducted a survey among several graduate students from the department of Anthropology. We divided them into 2 groups, and train one group how to use this program as opposed to the other group.

| No. | Time used to get familiar with the program |
|---|---|
| 1 | 57 minutes |
| 2 | 53 minutes |
| 3 | 61 minutes |
| 4 | 51 minutes |
| 5 | 64 minutes |
| 6 | 69 minutes |
| 7 | 77 minutes |
| Avg. | 61.7 minutes |

Table 4: Time used of the untrained group

| No. | Time used to get familiar with the program |
|---|---|
| 1 | 31 minutes |
| 2 | 34 minutes |
| 3 | 39 minutes |
| 4 | 28 minutes |
| 5 | 25 minutes |
| 6 | 33 minutes |
| 7 | 22 minutes |
| Avg. | 30.3 minutes |

Table 5: Time used of the trained group

We simply let the untrained group to figure out how the program works, and thought exploration, they did manage to discover most of it.

| No. | Rating (1-5) |
|---|---|
| 1 | 3 |
| 2 | 3 |
| 3 | 2 |
| 4 | 3 |
| 5 | 2 |
| 6 | 2 |
| 7 | 2 |
| Avg. | 2.4 |

Table 6: Overall rating of the untrained group

| No. | Rating (1-5) |
|---|---|
| 1 | 5 |
| 2 | 4 |
| 3 | 3 |
| 4 | 4 |
| 5 | 5 |
| 6 | 4 |
| 7 | 5 |
| Avg. | 4.3 |

Table 7: Overall rating of the trained group

We ask them to rate with the original implementation as the baseline. With proper training, the result is positive, while without guidance, the result is a little bit below our expectation.

## 4 Discussion

Visualizing uncertainty is a shortcoming of our visualization. Our current solution is to add a time period called unknown. Better visualization method can be applied to represent uncertainty. Also, there are more data available than pottery fragments in their open datasets. We plan to also visualize this information in the future.

## 5 Conclusion

We can see that we indeed provide a better way of presenting the data needed by research of anthropology. Through cautious investigation (Streamgraph, Treemap, Scatter plot), we make sure that proper techniques are used to build the relatively high quality visualization, like layering.

There are for sure space for improvement, and we believe that with more data and techniques (different views, linking) integrated, we can give anthropologists a better chance of discovering more.

## References


[1] Center for Comparative Archaeology, University of Pittsburgh. Settlement Patterns in the Chifeng Region, ISBN 978-1-877812-91-0, 2011.

[2] Alexander J. Martín. Comparative Archaeology Database, University of Pittsburgh. http://www.cadb.pitt.edu/chifeng/index.html.