Airbnb New User Bookings Visualization

CS519 Final Project Proposal

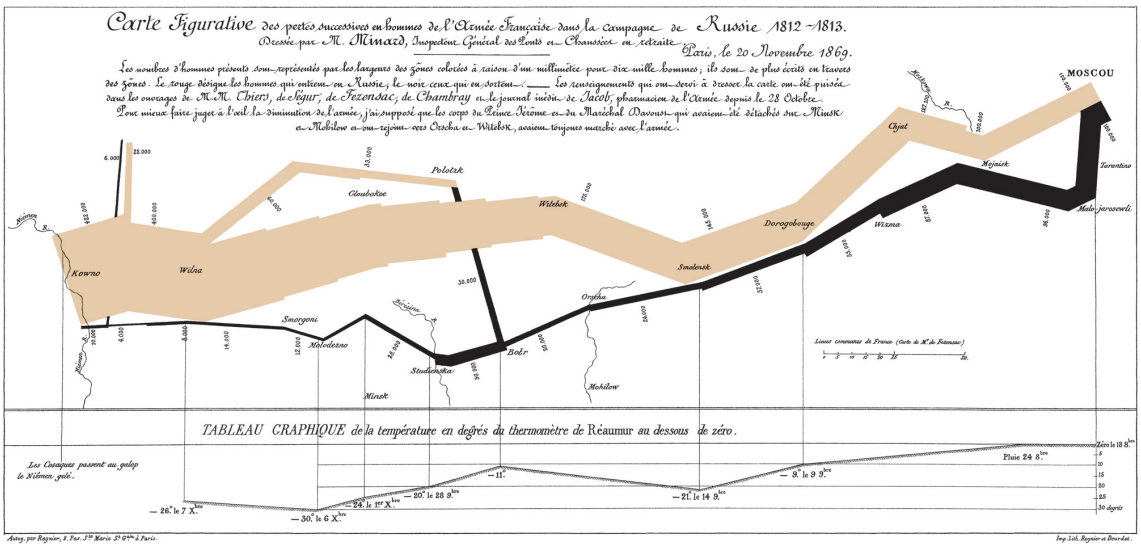
Xuan Wang, Wenyi Lu, Qian Wan

October 12, 2016

1 Introduction

We propose to make an online Hall-of-Fame visualization on Airbnb booking distribution, in which data comes from Kaggle competition. The visualization will show new users' multiple attributes with their relationships on where the new users will book their first travel experience. We choose Hall-of-Fame visualization because it is good at displaying multiple features information together. The attributes will include gender, age, sign-up method, language and so on.

This project is inspired by Hall-of-Fame Visualization on Napoleon's Invasion of Russia.The final image will look like



2 Team Member

Our team consist of 3 team members, listed as following:

1. Xuan Wang (xwang182)
2. Wenyi Lu (wl8)
3. Qian Wan (qianwan2)

3 Dataset

The dataset is from Airbnb New Users Booking in kaggle, it includes a list of users along with their demographics, web session records, and some summary statistics. We plan to use the following categories to show how will these information influence their final destination by an online Hall-of-Fame visualization.

Gender. It shows the gender of the account holder, and the gender can also be unknown because some people may not provide this information.

Age. It represents the age of the account holder.

Signup method. This part tell us how people sign up their Airbnb account, such as: basic, facebook, google, and so on.

Language. This category provide the language preference from the user.

Country destination. This category is the most important one for us, because our goal is to show the relationship between user's statistics and their first booking destination. And there are 11 possible outcomes of the destination country: 'US', 'FR', 'CA', 'GB', 'ES', 'IT', 'PT', 'NL','DE', 'AU', and 'other'.

4 Architecture & Environment

We intend to do the online 2D visualization using D3.js, JavaScript, HTML and CSS. It will be static visualization with no back-end server.

5 Proposed method

5.1 Problem Clarification

In this project, we want to get different attributes related with each other and build a highlight end-to-end style line to show the related data. The data is from Airbnb Company's rental records.

As we discuss in the Dataset part, there would be several attributes including sex, device used and many other information like user session in the dataset, how to combine them together and show them in one single visualization map becomes a big issue we need to solve here.

5.2 Proposed methodology & Project Deliverables

We are planning to tackle this problem on the baseline of Flow-based feature and End-to-end highlighting. In the visualization result, take nodes and links as input Data agent, while there would also be another version: instead of links, the input data contain flows, which have a single weight but multiple nodes in a chain.

The flows are easier to construct because our raw data are in a multi-attribute table(JSON version and .csv version with several attributes in each line), where each row corresponds to a flow. Moreover, flows retain more information than links. It knows where the flow is coming from at any given place. Thus the end-to-end highlighting is made possible.

New links between different attributes could be dynamically computed from this flow subset, and rendered in highlight in the future, and the endpoint should be consistent so the highlighted flows look visually consistent.

6 Verification & Validation

Since the dataset is new to us. We will evaluate the results by splitting the Kaggle dataset into a small dataset to do the stand-alone demo first, and then do the final overall dataset visualization.

References

- **Kaggle, Airbnb New User Bookings**
  https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings
- **Wikipedia, French invasion of Russia**
  https://en.wikipedia.org/wiki/French_invasion_of_Russia