FINAL REPORT

Visual Analytics

Contributors

Loren Resabala, Sachin Varule, Krishna Adhikari

Introduction: Domain problem characterization

Domain characterization is the first stage of visualization design, where visualization design denotes developing a digital system that allows users to find insight into data through visual representations and interaction methods.

We developed this user interface to be easy to use that allows the user to develop hypotheses and insights related to weather across the cities in USA.

What is the problem you want to solve?

We want to give more insights on:

1. How do weather and air temperature vary from city to city across the United States?
2. How has weather changes in time frame (Annual, 5 years, monthly) across the cities in USA.

Data/operation abstraction design:

We use the weather data set from Kaggle (access link in the Appendix) and to limit the scope of our project we decided to work only with the following csv files: City_attributes, Temperature, Wind_direction, and Wind_speed. Therefore, we have a dataset that includes hourly measurements for temperature, wind direction, and wind speed for selected US cities from 2013-2017.

Table below give us a snapshot of our dataset to have better understanding before data preparation.

Data Set	**Structure**	**Notes / Observations**
City_attributes	4 columns: City, Country, Latitude, Longitude	**City Column**: Only 36 cities listed **Country Column:** Unique value as United States
Temperature	37 columns: Datetime, and 36 columns representing city names.	Temperature measure is in **Kelvin (K)**
Wind_direction	37 columns: Datetime, and 36 columns representing city names.	Direction measure is in meteorological degrees[1]
Wind_speed	37 columns: Datetime, and 36 columns representing city names.	Speed is measure in **m/s** (meters per second)

[1] Meteorological degree wind direction is defined as the direction from which it originates. Wind direction is measured in degrees clockwise from due north.

How did you prepare your data?

The goal is to pre-process the data sets to format it **into a single data frame** to be used. We follow the process listed below:

1. Create a subset data frame that has the following variables: Datetime, City and Temperature
2. Convert Temperature from Kelvin (K) to Fahrenheit (°F). Kelvin to Fahrenheit conversion formula is **T(°F) = T(K) × 9/5 - 459.67**
3. Join this subset with the data sets of Wind (direction and speed) and City (latitude and longitude) resulting in a new data frame of 7 columns: Datetime, City, Temperature, Latitude, Longitude, Wind Direction, and Wind Speed
4. Because we have different time zones, use city longitude to convert time zone and have a rough approximation for cities.
5. Save this new subset in a csv file to be used in our Shiny app.

Encoding/Interaction design:

Based on the data, using a leaflet map to visualize our data would make the most sense. We also use a sidebar toggle which allows users to visualize multiple based on the input provided. User can choose 5 years, Yearly and Monthly. Along with the average period selected, users can also use a slider to select the day of the time to visualize individual time range.

Once the average period from the sidebar toggle is selected, users can choose which city they would like to explore more. The leaflet map allows users to visualize the average temperature in Fahrenheit for 36 cities. Once the city is selected from the map, users can see the name of the city selected, number of observations, and percent complete on the right side of the map. The number of observations represents the total number of datasets that are used in the visualization and the percent complete represents the total available data for the range selected. For example, if we selected January and percent complete is 93%, then it represents we are missing 7% of the data for the month of January. We feel that it's very important to let users know about data integrity.

We also decided to add a wind rose chart, which is a bar chart plotted using a polar grid. This wind rose chart allows users to understand the wind speed/pattern for the city within the average period selected.

We believe our visualization provides users a vast understanding of the data and simplifies the complex backend computation. Without having to know anything about the data users can make sense of what they are looking at.

Algorithmic design:

How does your systems work holistically (e.g., not sluggish)?

Different data filters (Time period, hours of the day) help to communicate results via interactive charts & visualizations.

Built-in browser deployment capabilities let you share your work easily with colleagues & Team.

User evaluation:

How would you test your system?

User experience is one of the major components of any design. One of the major items we discuss before starting the project was we wanted our design to be self-explanatory. Our users should understand the design with little or no explanation.

Within a limited time, we surveyed 8 colleagues and took feedback from them to make our system function flawlessly. Overall, the survey was very satisfying. We did get a couple of feedback about the location of the Map and its height and width. One of the surveyors mentioned that we add a temperature scale. After a brief discussion with the team, we decided to add the button to display the temperature scale which allows users to know which scale our application is using.

Future work

For future work we can integrate more datasets to this visualization like data related to pollution, deforestation happening in the cities etc.