Kaarina Tulleau
INFO 474

Final Project Overview

Design Overview

This visualization is a two-sided clustered bar chart, with one side displaying temperature-related data, and the other displaying precipitation-related data. The data is aggregated at the month level. Details on how the data was processed prior to the visualization can be found in the additional information section of this document. The visualization is intended to allow simultaneous comparison between different temperature values and precipitation values on a monthly basis. The visualization also allows for the identification of patterns in weather and precipitation, by comparing temperature and precipitation data across the x-axis. It is designed to answer the following questions: "What are temporal trends in temperature for different cities over the years 2014-2015?", "What are temporal trends in precipitation for different cities over the years 2014-2015?", and "Is there a pattern/correlation between temperature and precipitation in different cities over the years 2014-2015?".

The visualization provides several opportunities for user input, so that the user can fine-tune the data displayed to answer their specific questions. The user can select the location (city) and year for which to display data. They can then select what data (average minimum temperature, etc) they would like to display for each of the four columns on the bar chart. The chart title is dynamic to reflect the location and year selected, and the legend is updated whenever the user changes the data (columns) displayed. The column dropdowns also indicate which columns on the bar chart they correspond to, in order to minimize confusion.

User Tasks Supported

The visualization supports the following user tasks, among others:
- As a user, I want to be able to filter the visualization based on location so that I can decide which city's data to display.
- As a user, I want to be able to filter the visualization based on year so that I can view the data for a specific timeframe (year).
- As a user, I want to be able to select specific data (columns) to display on the temperature side of the bar chart, so that I can compare them and identify any patterns/trends.

- As a user, I want to be able to decide which temperature and precipitation values to display, so that I can identify any patterns/correlations in temperature and precipitation levels.
- As a user, I want to have easy access to the data provenance for the visualization so that I can be aware of the source and reliability of the data, since I am using it to make decisions.
- As a user, I want to have easy access to a definition of the data (columns) available for display, so that I can correctly interpret the visualization.

User Interface
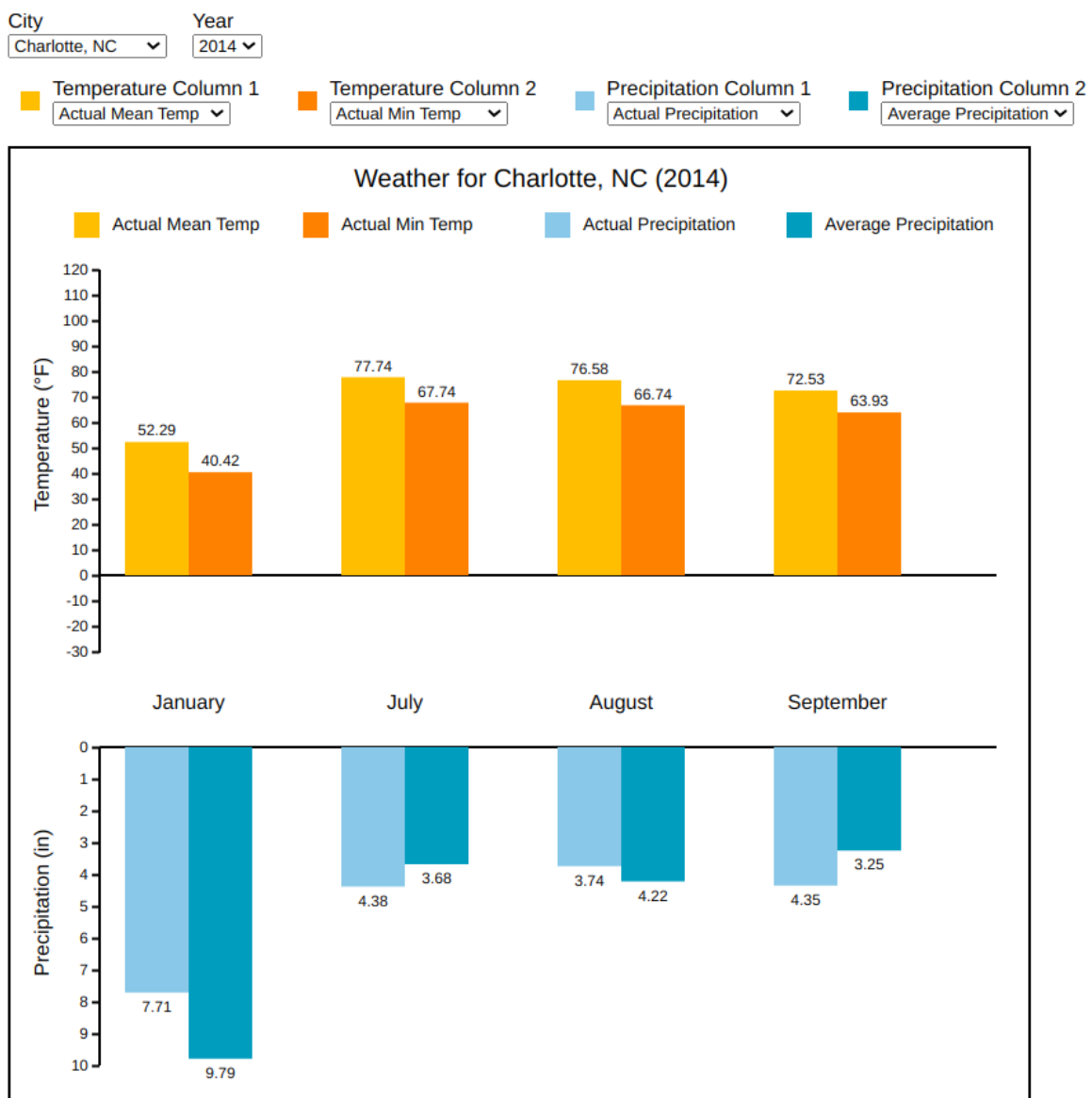
Additional Information

This section contains some notes on data processing and color selection for the visualization.

The data processing for this project was done in python, and the processing.ipynb file is included in the zip file. The data processing followed these general steps:

1. Load CVSs into pandas dataframes.
2. In each dataframe, create month & year columns for later aggregation.
3. In each dataframe, remove the record_min_year and record_max_year columns prior to aggregation.
4. Group/aggregate the dataframes on year, month. The columns were aggregated as follows:
   a. Mean: actual_mean_temp, actual_min_temp, actual_max_temp, average_min_temp, average_max_temp
   b. Min: record_min_temp
   c. Max: record_max_temp, record_precipitation
   d. Sum: actual_precipitation, average_precipitation
5. Add the city/state information to each dataframe.
6. Vertically join all dataframes together.
7. Export to final.csv, which is used in the visualization.

An orange/blue color scheme was chosen for two reasons. One, temperature is often associated with heat, which typically brings "warm" colors (such as red and orange) to mind, while precipitation involves rain and snow, which brings to mind "cold" colors (such as white and blue). Two, the orange/blue color scheme is one of the best for color blindness, with most kinds of color-blind people able to perceive differences between the two colors. The shades of blue/orange are also distinct enough that someone could distinguish between them in greyscale.