Tammy Zhang

INFO 4310

Homework 1

Stories

In these visualizations, I aimed to explore the timing with which trees are planted in San Francisco on multiple scales - monthly, seasonal, and annual, with a focus on seasonal faceting. The story conveyed by the data includes the following:

- While it is possible to observe some successive rows of trees planted in the same season on the map, potentially due to targeted planting projects coordinated at the same time by the city, there doesn't appear to be any obvious geographic boundaries as to how tree planting is timed seasonally. Tree planting occurs year-round across many neighborhoods. (Map Component)
- Tree planting peaks in late winter/early spring and is significantly decreased during fall. This aligns with some local gardening advice stating to plant trees in late winter (1, 2). (Histogram and Time Series Components)
- Seasonal tree planting fluctuates on a yearly basis, sometimes on a dramatic scale, but across all seasons, the last few decades have seen significant increases in tree planting numbers compared to 1950 - 1990. (Time Series Component)

Visual Encodings

I included a map of trees color coded according to the season they were planted, a histogram showing the monthly distribution of tree plantings, and four time series plots showing how many trees were planted per season each year since 1950. For the map, the visual encoding

was each tree being represented by a circle that could take one of four fill colors - green, yellow, red, or blue, depending on the season it was planted. The horizontal and vertical positions of the circles correspond to longitude and latitude, respectively. The same colors were used in the histogram - which used bars - and in the time series, which each used a line. For the histogram, one bar was made per month; their horizontal position was given by the order of the months (an ordinal category) while their vertical position/length (aka height of the bar) indicated tree planting counts. For the time series lines, horizontal position represented time (year) while vertical position indicated tree planting counts.

Data Processing

In order to process the data, I took all entries from the filtered dataset that had a valid PlantDate value (not nan), then used Python's datetime library to parse out the months and years. I assigned each tree a planted season based on March, April, May = spring; June, July, August = summer; September, October, November = fall; December, January, February = winter. I used pandas' value_counts() function to create datasets that binned the counts by month and year as well for use in the histogram and time series plots. The code can be found in "Data Processing.ipynb".

Design Rationale:

The formats were chosen in order to illustrate one particular aspect of the data - seasonality - across different scales. The map allows for a broad geographic view; the data was rendered at a tree-level instead of aggregating in order to see if any particular small clusters or patterns of similar seasons would emerge. The tradeoff with this approach is how parts of the map can become visually crowded. The histogram was chosen to show planting distribution over

the course of a year (one full set of seasons). The time series were chosen to show planting distribution over the course of multiple years. I faceted the seasons into four separate plots since combining them into one became crowded (and aligning seasonal values horizontally based on their year hides the passing of time between seasons), but the tradeoff is that this may make direct comparison between seasons more difficult. I made sure to use a y-axis scaled to the same global maximum value in order to not distort comparisons.

Another challenging design decision was the color palette - I chose a categorical, discrete color scale with four values picked for their visual contrast with each other and associations with the season (e.g. blue for winter, the colder months, and green for spring, when leaves return). I considered applying colors on the map and histogram in a sequential fashion, because seasons do have somewhat of a natural order - but found that having shades of one color becomes difficult to distinguish, particularly on a map with small dots where many may be overlapping. I used the same colors to encode the same seasons across all visual elements for consistency; I considered including an explicit color legend, but thought that this might decrease the data-to-ink ratio and add information that can be understood from the histogram and time series.