

Edlab webpage: <http://www-edlab.cs.umass.edu/~yuanh/590v/hw2/index.html>

Interactive Visualizations for the Facebook posts dataset and Wine dataset

The facebook dataset is public available for research. The details are described in [Moro et al., 2016].

The wine dataset is public available for research. The details are described in [Cortez et al., 2009].

Data Preprocess

For Facebook data:

Use pandas library in python to first fill the empty attributes with 0. Change the “Category” value from 1,2,3 to “inspiration”, “action” and “product”. Change the “Post Weekday” from numbers to strings “Mon”, “Tue”, etc. Save the clean data in a csv file.

Save all the numerical feature names and categorical feature names in a dictionary. Later we will use this metadata information to determine the axes in different visualizations.

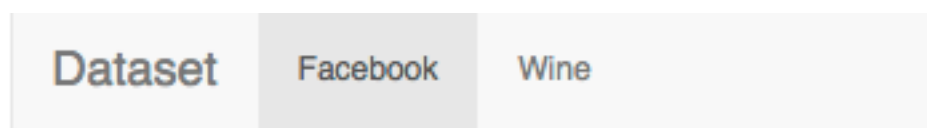
For Wine data:

Use pandas library in python to combine the red wine and white wine datasets. Add the type “red” and “white” as a new feature in the data table. Fill in all empty attributes with 0. Save the new clean data in a csv file.

Save all the numerical feature names and categorical feature names in a dictionary. Later we will use this metadata information to determine the axes in different visualizations.

Visualizations

The webpage has a navigation header which allows us to select the visualizations for the Facebook or Wine dataset.

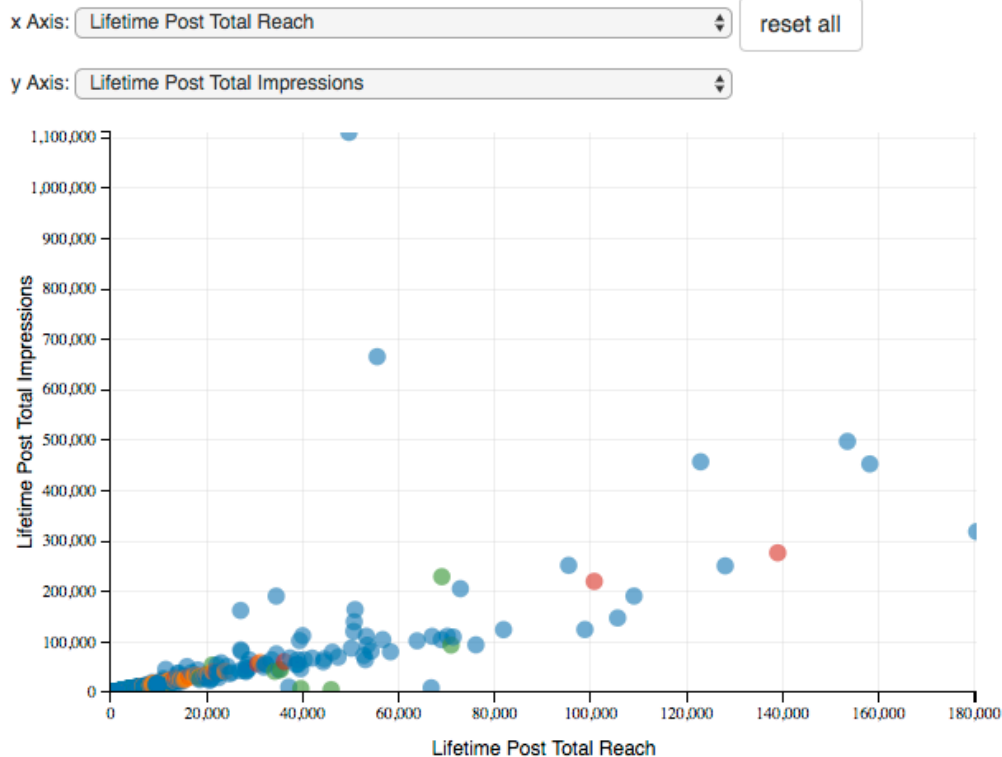


Facebook

Scatter Plot

Plot two different numerical features as x and y axis in a scatter plot. Use different color to represent different types. Each point can be selected. The x and y axis can be changed by user. The color legend is controlled by the pie chart next to it. The scatter plot also supports probing.

Lifetime Post Total Impressions By Lifetime Post Total Reach



Color Legend: The bubble chart shares the same color legend with the pie chart to the right.

Description: The Lifetime Total Reach and Lifetime total Impression of the facebook posts have an approximately linear correlation with some outliers. The posts with high Total Reach and Total Impressions are all photos and videos. Status and links have less reach and impressions.

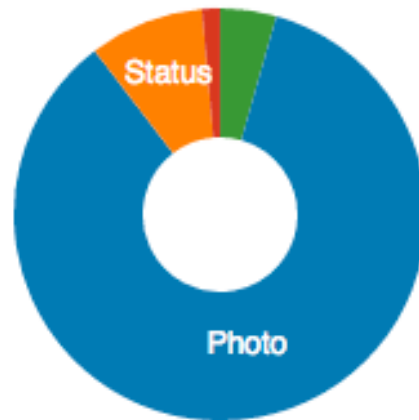
Pie Chart

Plot the count of posts with different categorical features (such as type, category, post weekday, paid/unpaid) in a pie chart. The color is the same with the scatter plot and the attribute attached to color can be changed in the dropdown menu.

By Type

color:

- Link
- Photo
- Status
- Video



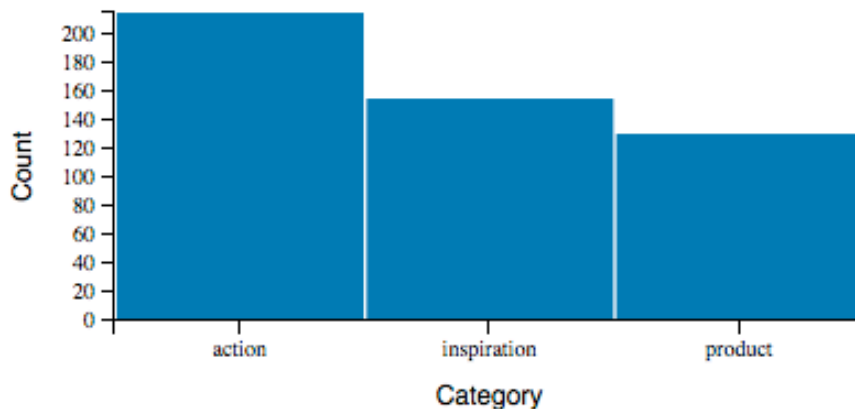
Description: The pie chart shows the count and percentage of different types, categories, post weekday, paid/unpaid among all the posts. Most of the posts are photos.

Each slice of the pie chart supports probing (will show the count in each slice and the percentage) and selecting.

Bar Chart (Histogram)

By Category

x Axis:



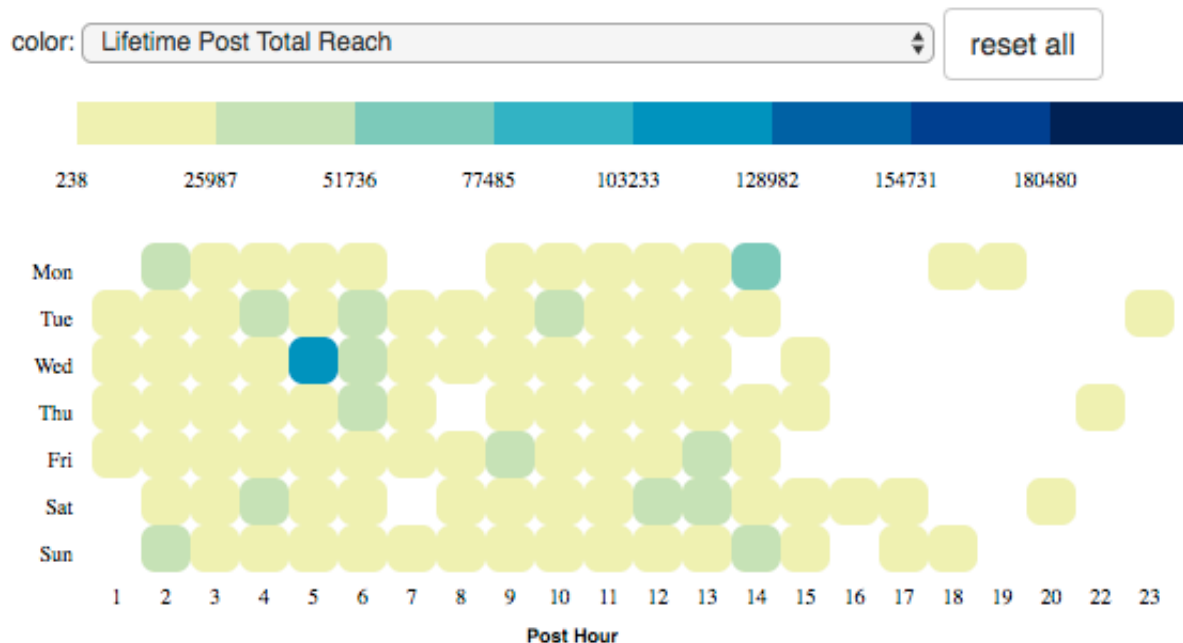
Description: Histogram of different categories, types, post weekday, hour, month, paid/unpaid of all the posts. Select one category, and we can find in the pie chart: 97% of the 'Inspiration' category are photos, while in 'product' category, there are 29% status and 70% photos.

Plot the histogram of posts with different categorical features (such as type, category, post hour, post month, post weekday, paid/unpaid) in a bar chart. Color doesn't have any meaning. Each bar supports probing and selecting.

Heatmap

The heat map of different numerical features (Lifetime Total Reach, etc.) distributed at different post hour and weekday is shown. At each box the color represents the average value for all posts posted at a certain hour and weekday. Probing and selecting is supported.

Lifetime Post Total Reach



Description: The heatmap shows the distribution at different post hour and weekday. The color can represent different measures for the popularity of the posts. The most popular posts are posted in the morning and around noon.

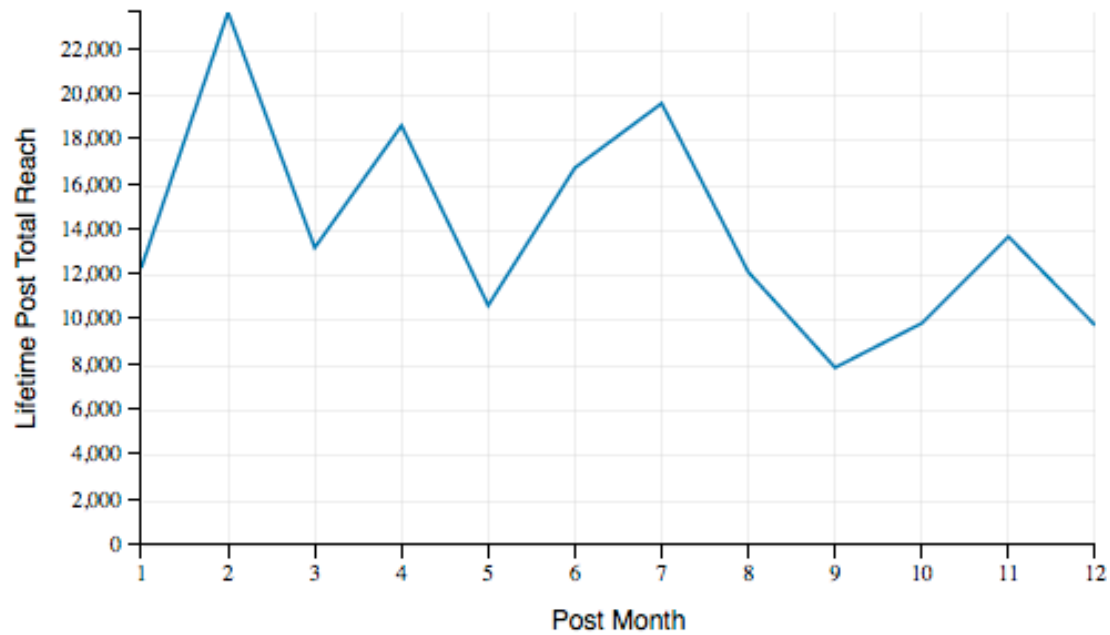
Line Chart

Different numerical features (Lifetime Total Reach, etc.) as a function of post hour and post month is plotted as a line chart. The line chart supports brushing which allows to select multiple points at the same time but does not support probing.

Lifetime Post Total Reach By Post Month

x Axis:

y Axis:



Description: The line chart shows how the measure of popularity depends on post Month and post hour.

Wine

There are only two categorical features in the wine data, which are the type(red/white) and quality. Therefore, for the wine dataset, I use the same visualization code but only keep the scatter plot, the pie chart and the histogram.

Scatter Plot

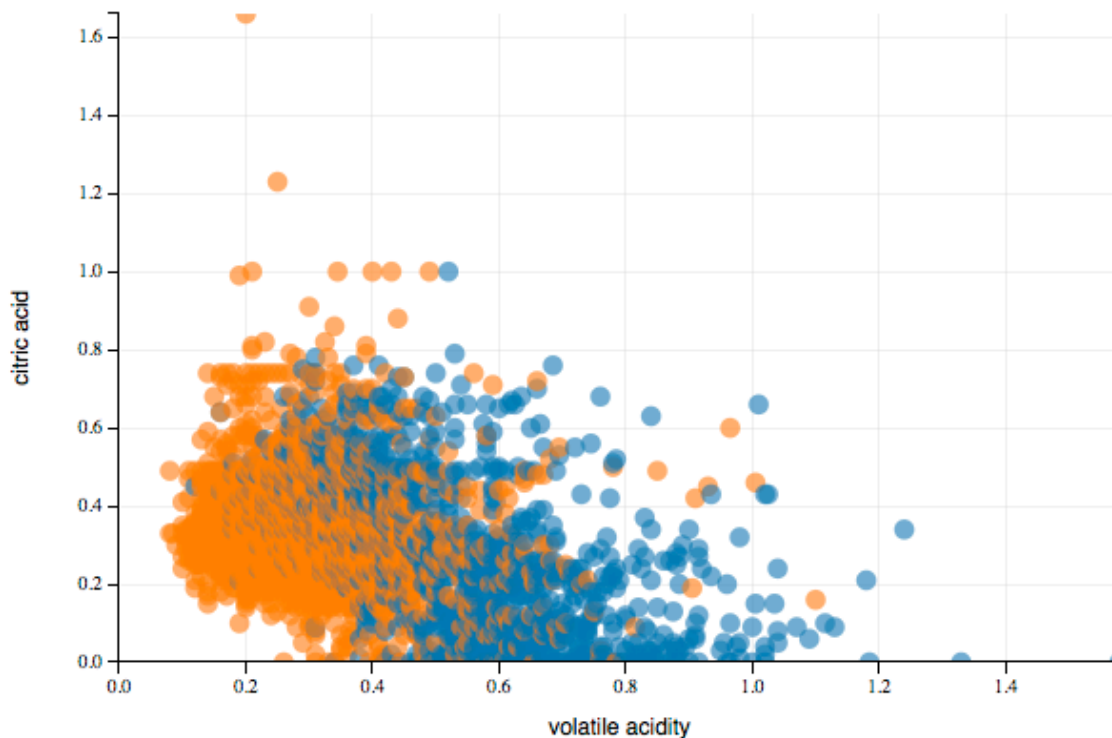
Plot two different numerical features as x and y axis in a scatter plot. Use different color to represent different types(red/white wine). Each point can be selected. The x and y axis can be changed by user. The color legend is controlled by the pie chart next to it. The scatter plot also supports probing.

citric acid By volatile acidity

x Axis: volatile acidity

reset all

y Axis: citric acid



Color Legend: The bubble chart shares the same color legend with the pie chart to the right.

Description: The white and red wines clearly form in to two clusters. White wines have lower volatile acidity, and higher citric acid. Select different properties to see how they differ for white and red wines.

Pie Chart

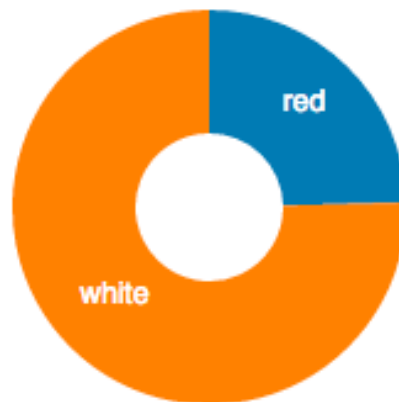
Plot the count of posts with different categorical features (type and quality) in a pie chart. The color is the same with the scatter plot and the attribute attached to color can be changed in the dropdown menu. Each slice of the pie chart supports probing (will show the count in each slice and the percentage) and selecting.

By Type

color: Type

reset all

red
white



Description: There are 75% white wine and 25% red wines in the data set. You can also use quality as the color attribute and see the scatter plot for wines of different qualities. There are no clear clusters and different properties vary case by case.

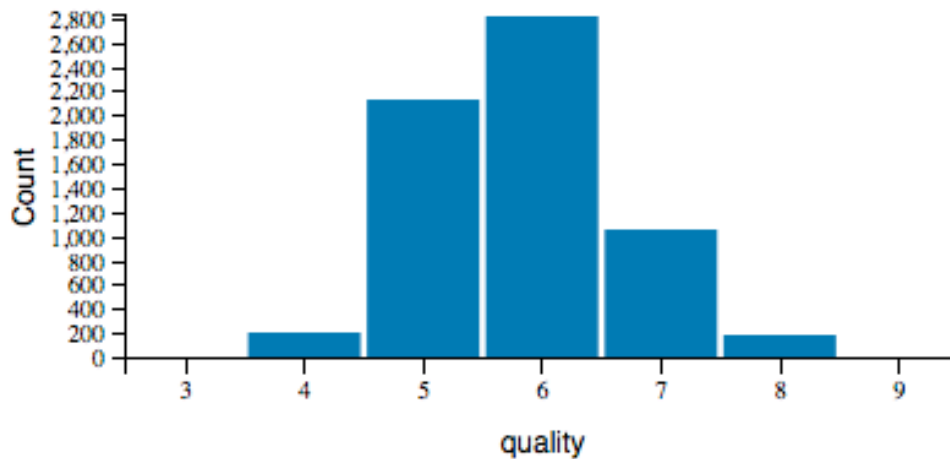
Bar Chart (Histogram)

Plot the histogram of posts with different categorical features (type and quality) in a bar chart. Color doesn't have any meaning. Each bar supports probing and selecting.

By quality

x Axis: quality ↕

reset all



Description: The histogram for wines with different qualities. You can select one quality level to see different properties like volatile acidity and citric acid from the scatter plot. If we only look at the good wines (quality 7 and above), the volatile acidity and citric acid of white wines are located in a compact cluster. While the citric acid of good red wines seem to form into two clusters.

References

[Moro et al., 2016] S. Moro, P. Rita and B. Vala. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. Journal of Business Research, Elsevier, In press. Available at: <http://dx.doi.org/10.1016/j.jbusres.2016.02.010>

[Cortez et al., 2009] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.