

Homework 1: Data Analysis Basics

Problem

Obtain the CSV (comma separated variable) file containing the counts of bicycles crossing the Fremont Bridge since 2012 (as described in <https://data.seattle.gov/Transportation/Fremont-Bridge-Hourly-Bicycle-Counts-by-Month-Octo/65db-xm6k>). Create a project directory with subdirectories for data and analysis, and create a README file. Download the data from <https://data.seattle.gov/api/views/65db-xm6k/rows.csv?accessType=DOWNLOAD> and put it in the data directory. Create a Jupyter notebook to analyze these data. In the notebook, complete the following:

1. Read the CSV file into a pandas dataframe
2. Add columns to the dataframe containing
 - a. the total (East + West) bicycle count
 - b. the hour of the day
 - c. the year
3. Create a dataframe with the subset of data from the year 2016
4. Use pandas + matplotlib to plot the counts by hour (i.e. hour of the day on the x-axis, total daily counts on the y-axis)
5. Use pandas to determine what is (on average) the busiest hour of the day

Note that we fully expect this analysis to cover some unfamiliar ground, and require teaching yourself a bit about Python and/or the Pandas package. Part of the intent of this assignment is to give you practice seeking help via the web, which (in our experience) is an essential part of using any data science tool. For example, if you type a question about Pandas into Google, you'll often find an existing answer to your question or something similar on the website StackOverflow.

A couple other online resources that might be helpful as you work through this:

- [The Python Data Science Handbook](#) (free online) has a chapter devoted to Pandas
- [Jake's Jupyter Workflow Video Series](#) shows some examples of working with this particular dataset in a Jupyter notebook.