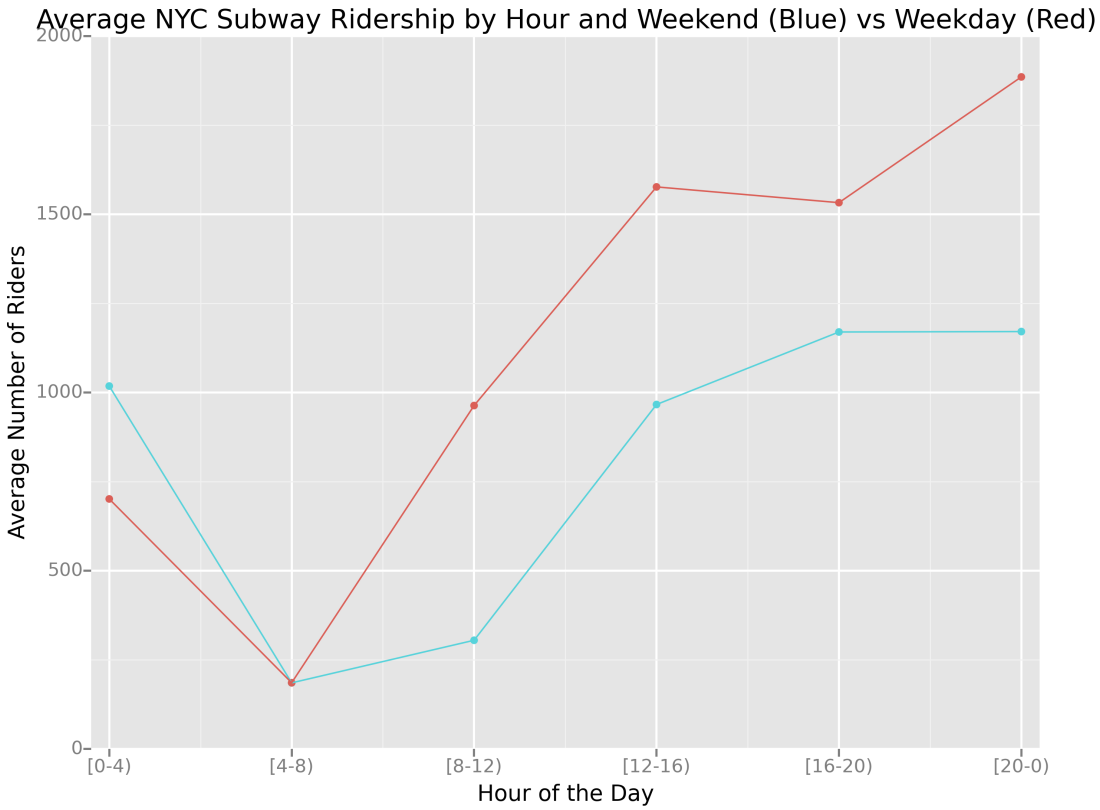
Examples

Plots

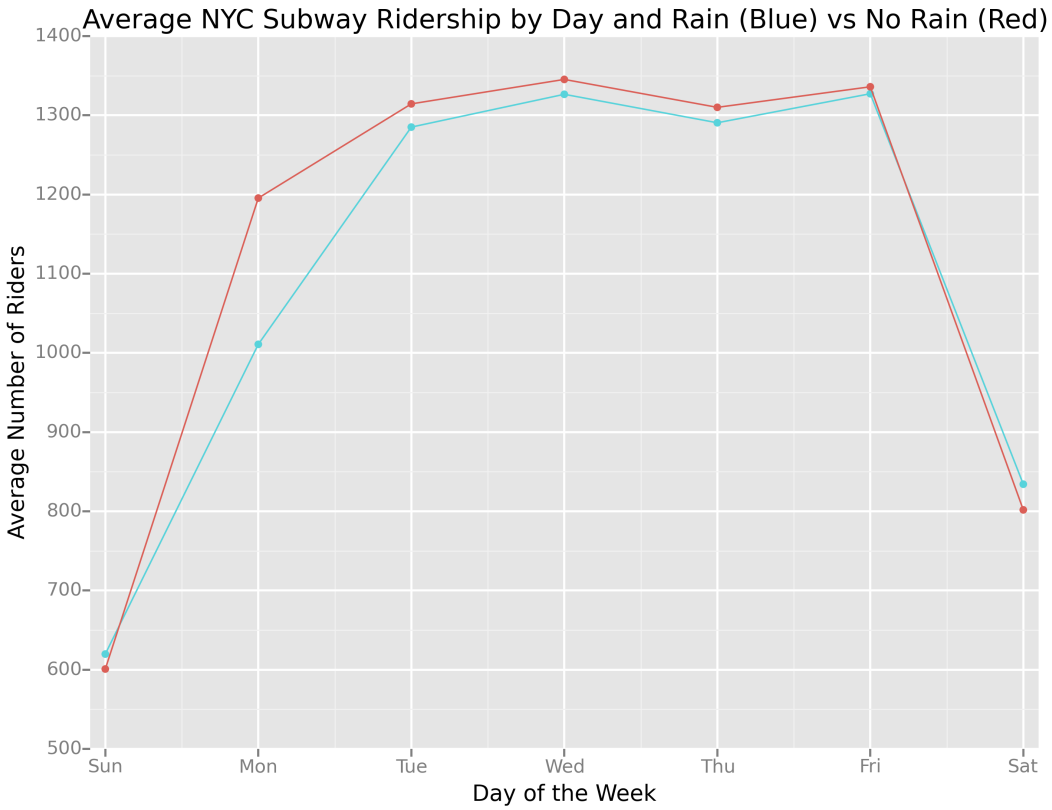
Ridership by Hour and Day Type

This figure shows subway ridership by time of day for weekdays and the weekend. First, I combined each of the entries into four-hour bins and then averaged them to help account for the variability in the hourly readings (mainly due to when the data happened to be recorded). I then split the binned ENTRIESn_hourly data into two groups: those recorded on a weekday (Monday-Friday) and those recorded on a weekend (Saturday-Sunday). It is important to note that these values correspond to the hour when the readings were taken, not necessarily when they actually occurred. The entries were mostly read every four hours, so some of the entries in a reported value could have been from 3:59 hours earlier. Also note that the labels indicate half-open intervals for each bin (e.g., "[4–8)" indicates 4:00–7:59).



Ridership by Day and Rain Status

This figure shows subway ridership by day of the week and rain status. First, I created a new field to indicate the day of the week (as we did earlier in one of the problem sets). I then grouped all of the ENTRIESn_hourly data by day of the week and rain status (i.e., if it rained or not that day; this was done using the “rain” field we created earlier).



Feature Selection

Using StatsModels OLS, the coefficients are:

```
rain            10.046302
Hour            62.239983
mintempi       -15.976535
maxtempi         3.794069
meanwindspdi    33.752174
fog            274.059647
precipi        -81.110211
```