The Turnstile weather dataset includes data regarding NYC subway use during the month of May 2011. Number of entries and exits are aggregated according to different sensors located within different Subway stations. The dataset also includes details regarding weather conditions. This makes possible to assess whether Subway use is affected by weather conditions and to what extent.

Plotting a histogram of hourly entries during rainy and not rainy days it is possible to clearly see how none of the two is distributed following a normal distribution.

References used

<http://pandas.pydata.org/pandas-docs/stable/>

<http://scikit-learn.org/stable/modules/sgd.html>

<http://docs.ggplot2.org/0.9.3.1/facet_wrap.html>

<http://pandas.pydata.org/pandas-docs/stable/advanced.html>

<http://pandas.pydata.org/pandas-docs/stable/indexing.html>