

Data Visualization

Part 1

Jonathan Arriaga, PhD

Index

- Bar Plots for Categorical Data
- Scatter Plots
- Scatter Matrix
- Time Series

Bar Plots for Categorical Data

- Very effective method to represent categorical data.
- Each bar represents the occurrences of an event.

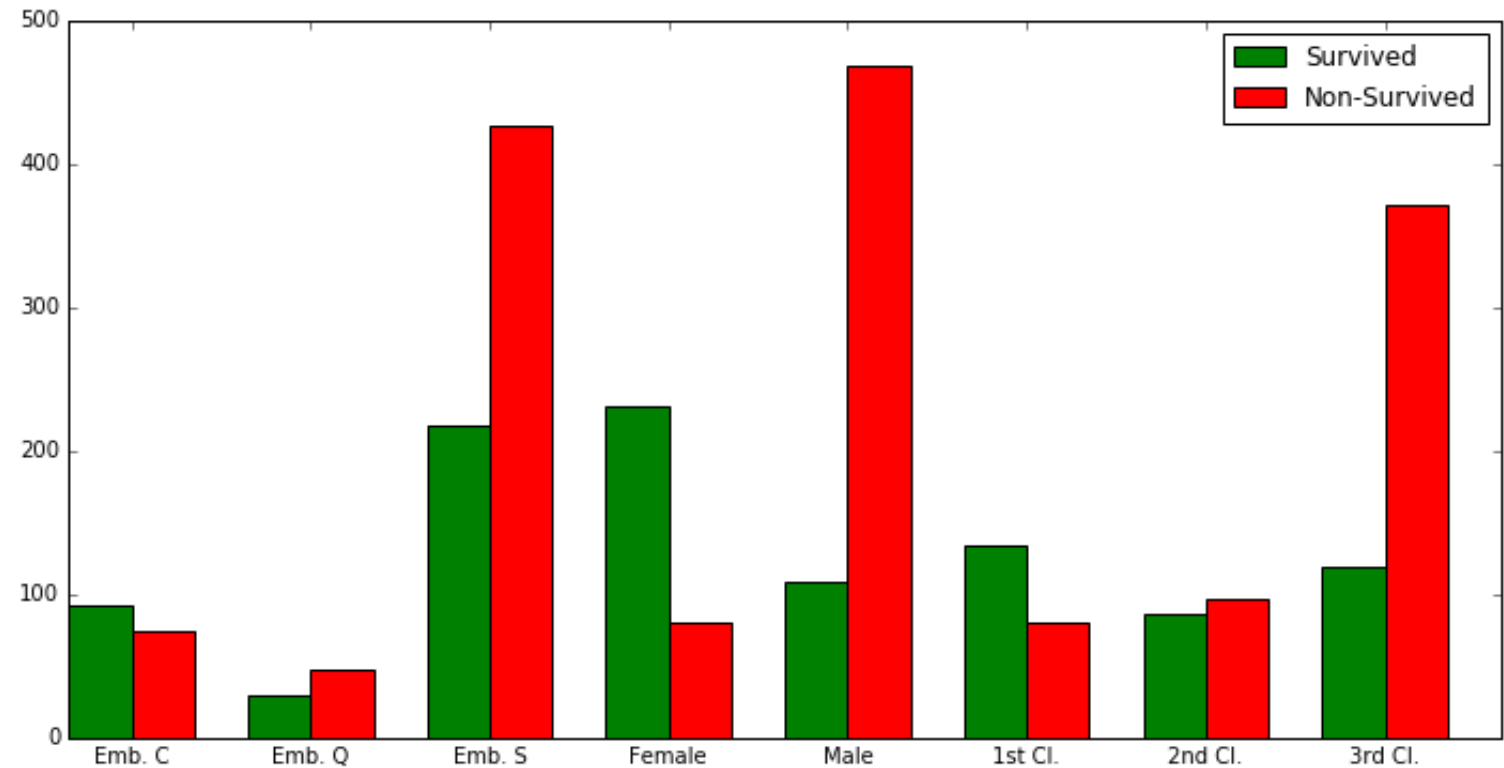
Dataset: Titanic

Features used:

- embarked
- sex
- pclass

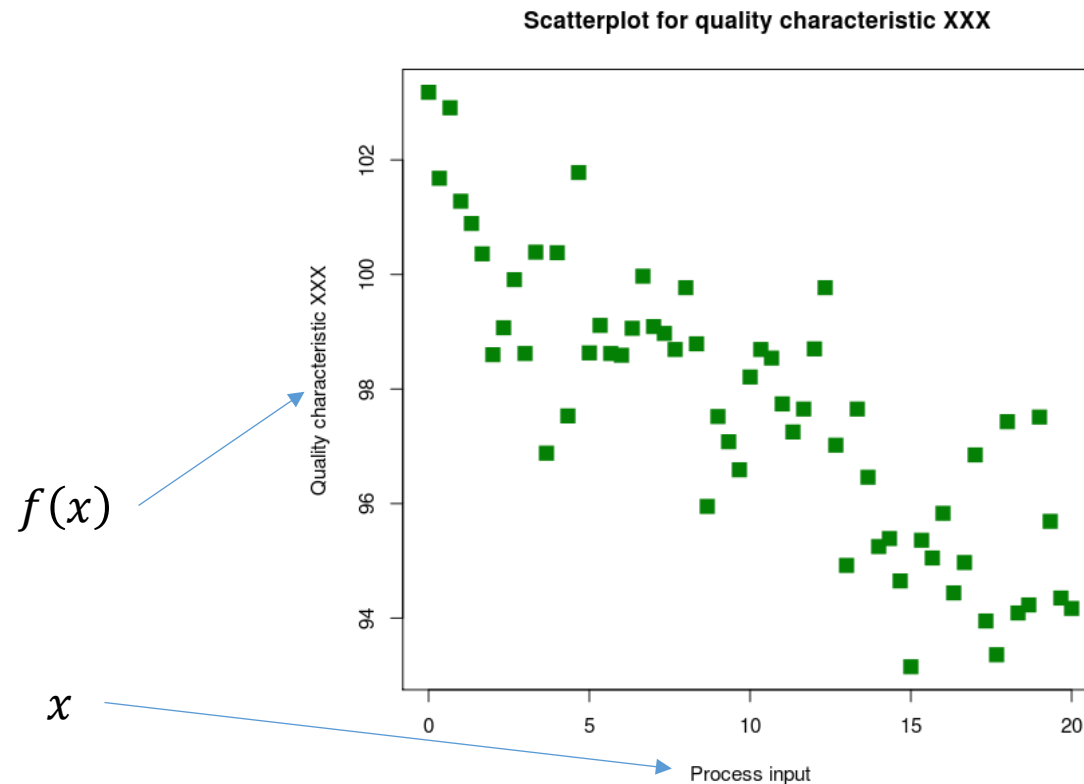
Source:

<https://www.kaggle.com/c/titanic/data>



Scatter Plot

- Typically used to display the relationship between a control variable and a dependent variable.
- Points can be displayed in different colors, shapes, and sizes to visualize additional dimensions.



Scatter Plot Example

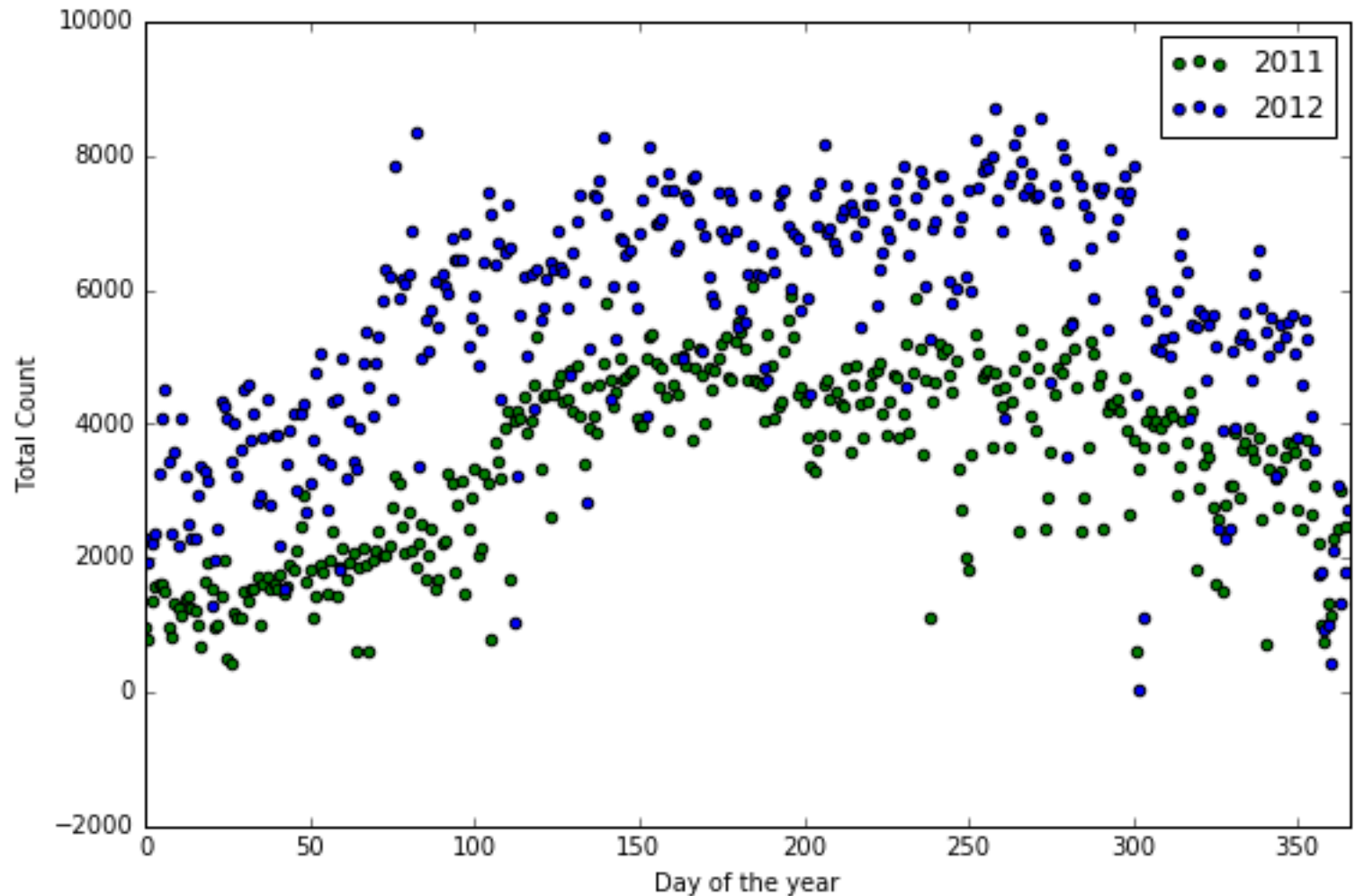
Dataset: Bike Sharing

Features used:

- yr
- cnt

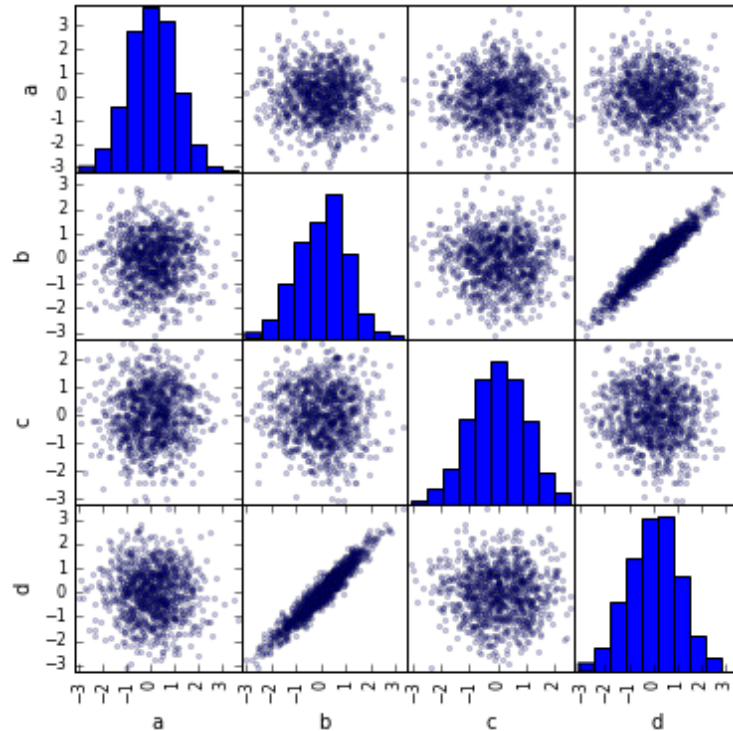
Source:

<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>



Scatter Matrix

- Visualize the relationships on multiple variables by pairs.
- Determine if a linear correlation exist between a pair of variables.
- The distribution for individual variables is depicted in the diagonal, it can be a histogram, density plot, boxplot, or other plots depending on the visualization package.



Scatter Matrix Example

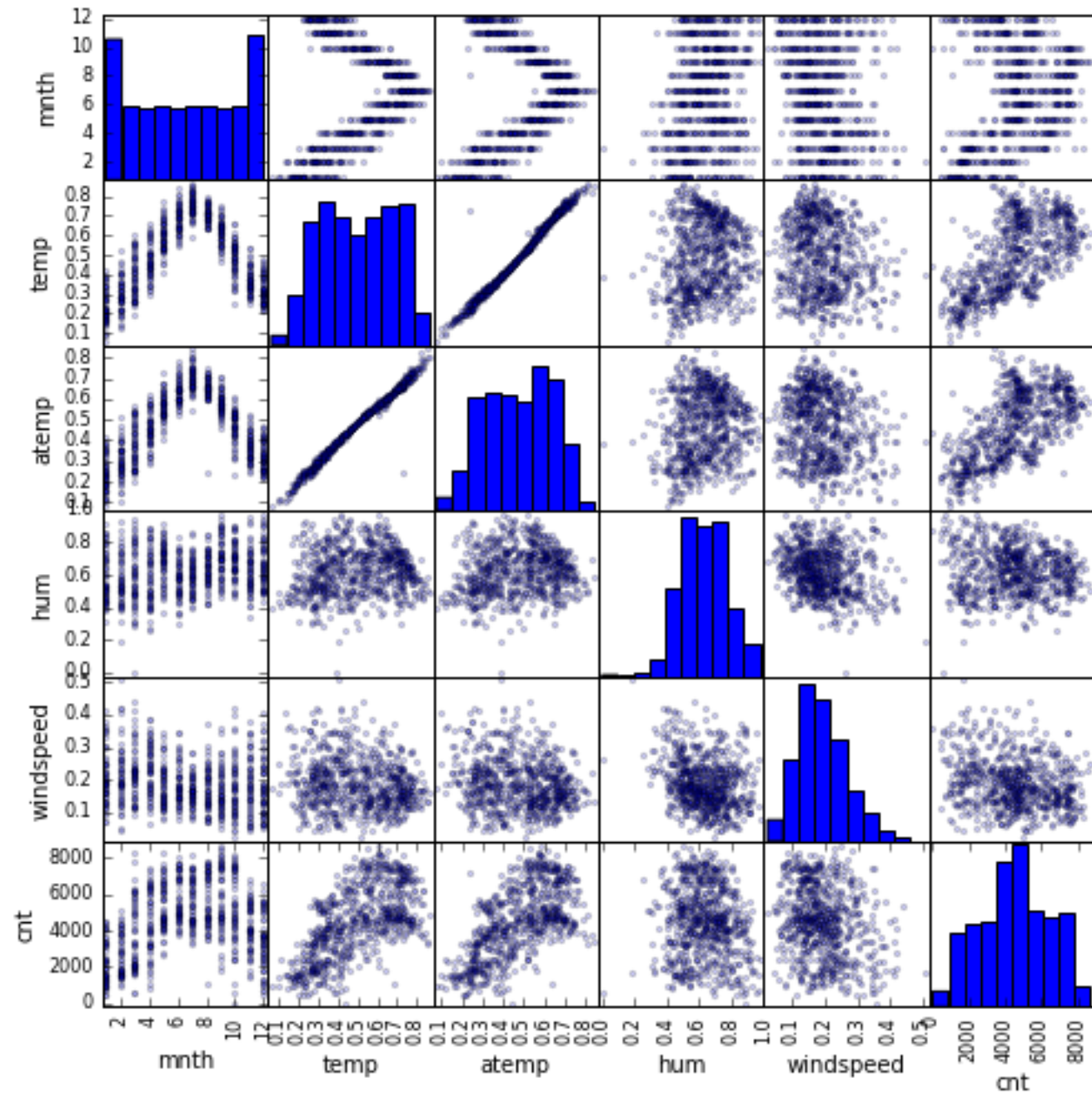
Dataset: Bike Sharing

Features used:

- mnth
- temp
- atemp
- hum
- windspeed
- Cnt

Source:

<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>



Time Series

- Plot sequential or time-related data.
- Operations on a moving/rolling window are applied to extract features on the data.

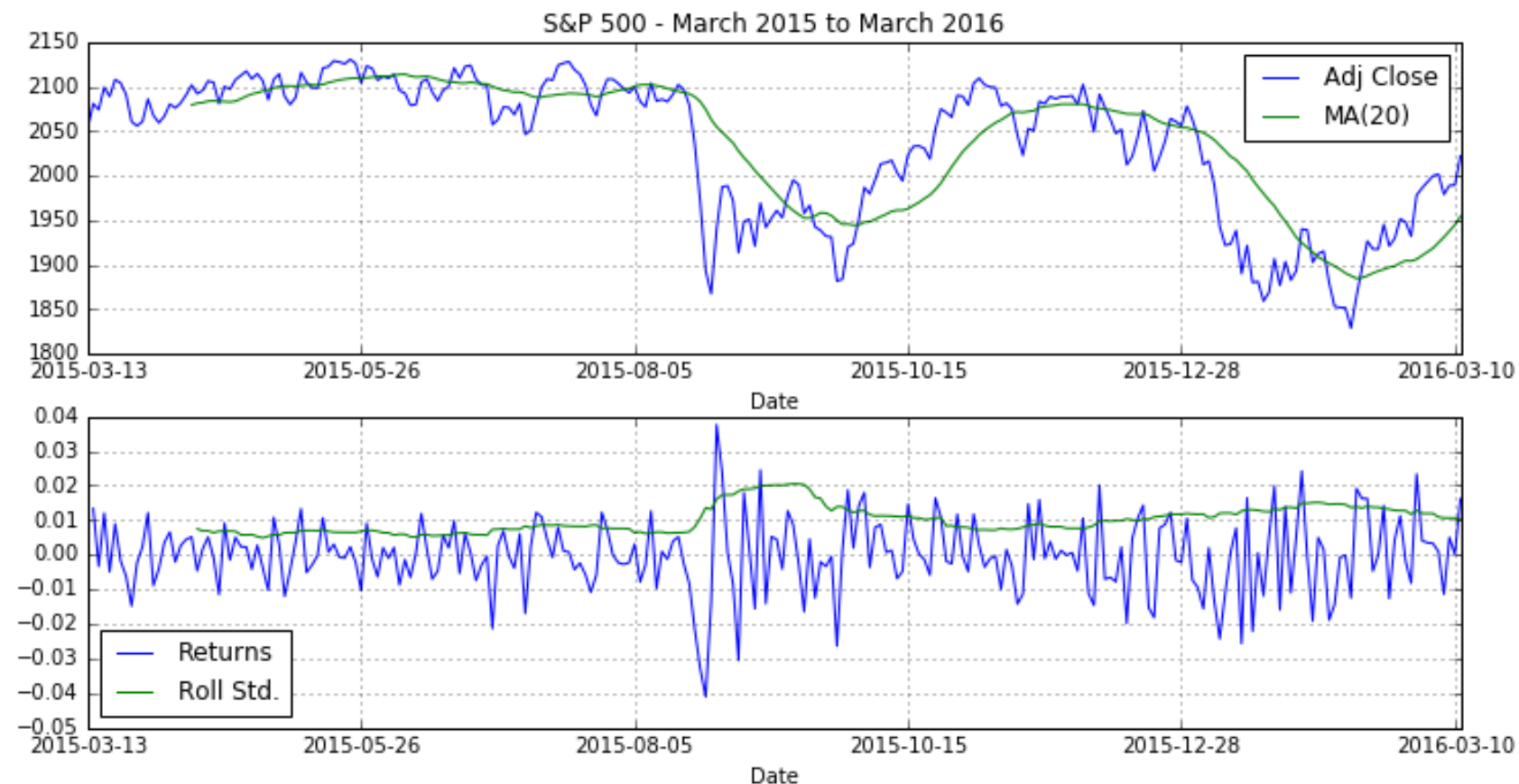
Dataset: S&P 500 Index from
March 2015 to March 2016

Features:

- Date
- Adj Close

Source:

<http://finance.yahoo.com/q/hp?s=%5EGSPC+Historical+Prices>



Datasets used for the examples in this lecture

- **Titanic Dataset**

<https://www.kaggle.com/c/titanic/data>

- **Bike Sharing Dataset**

<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>

- **S&P500 Data**

<http://finance.yahoo.com/q/hp?s=%5EGSPC+Historical+Prices>