Visualizaci?n de datos

January 31, 2018

```
In [4]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import matplotlib.mlab as mlab
        import math
        import seaborn as sns
        from IPython.display import Image
        %matplotlib inline
```

0.1 Visualización de Datos

http://www.slideshare.net/idigdata/data-visualization-best-practices-2013

```
In [5]: Image(url="http://istc-bigdata.org/wp-content/uploads/2014/09/ISTC-Blog-Automating-Visu
```

```
Out[5]: <IPython.core.display.Image object>
```

0.1.1 Herramientas para la visualización de datos en Python (De Python Awesome List: https://github.com/vinta/awesome-python/blob/master/README.md#data-visualization)

- matplotlib - A Python 2D plotting library.
- bokeh - Interactive Web Plotting for Python.
- ggplot - Same API as ggplot2 for R.
- plotly - Collaborative web plotting for Python and matplotlib.
- pygal - A Python SVG Charts Creator.
- orange - Data mining, data visualization, analysis and machine learning through visual programming or Python scripting.
- pygraphviz - Python interface to Graphviz.
- PyQtGraph - Interactive and realtime 2D/3D/Image plotting and science/engineering widgets.
- SnakeViz - A browser based graphical viewer for the output of Python's cProfile module.
- vincent - A Python to Vega translator.
- VisPy - High-performance scientific visualization based on OpenGL.

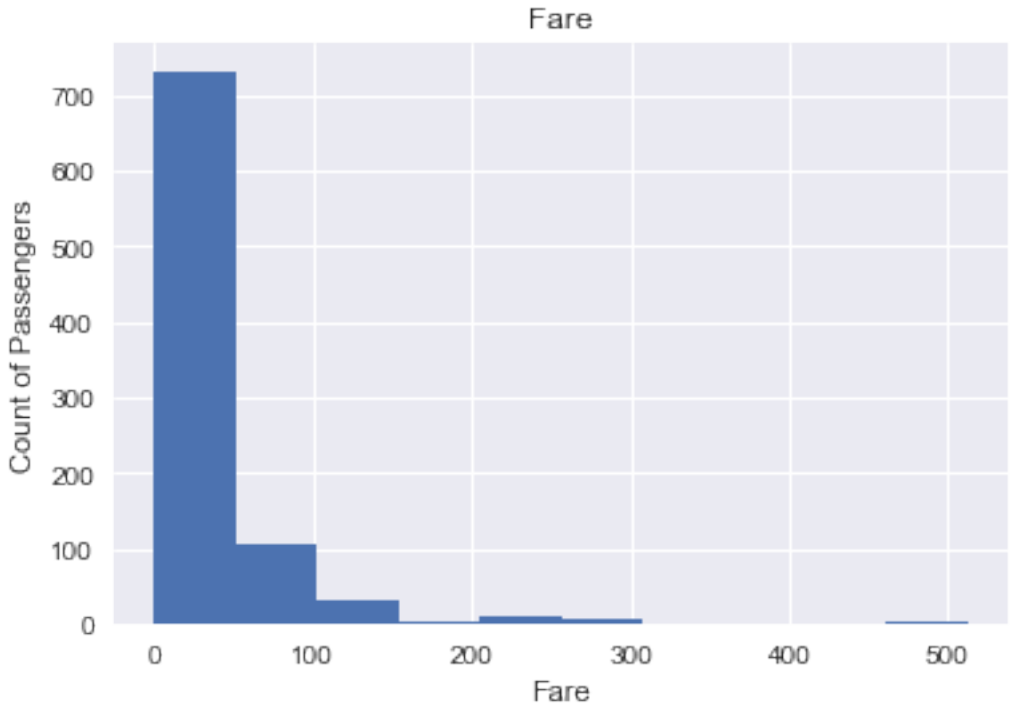
0.1.2 The grammar of graphics

http://ggplot2.org/resources/2007-past-present-future.pdf

In [7]:
```
#Cargamos los datos
df = pd.read_csv("../train.csv")
```

In [8]:
```
print("skewness:", df["Fare"].skew())
print("kurtosis:", df["Fare"].kurtosis())
fig = plt.figure()
ax = fig.add_subplot(111)
x = np.linspace(df['Fare'].min(),df['Fare'].max(),1)
ax.hist(df['Fare'], bins = 10, range = (df['Fare'].min(),df['Fare'].max()))
plt.title('Fare')
plt.xlabel('Fare')
plt.ylabel('Count of Passengers')
plt.show()
```

```
skewness: 4.78731651967
kurtosis: 33.3981408809
```



In [10]:
```
from bokeh.charts import Bar, output_file, show
from bokeh.sampledata.autompg import autompg as df

p = Bar(df, 'cyl', values='mpg', title="Total MPG by CYL", notebook=True)
output_file("bar.html")
show(p)
```

0.1.3 Herramientas externas para la visualización de datos

- Tableau
- raw.io (http://app.raw.densitydesign.org/)
- easil.ly

0.1.4 D3.js

0.1.5 Actividad: Elige cualquiera de los métodos que vimos, y visualiza el resultado del análisis que hicimos durante el workshop

0.1.6 Recursos:

- http://www.cs171.org/2016/index.html
- https://github.com/rodowi/geo-is-osm
- http://blog.hubspot.com/marketing/excel-graph-tricks-list