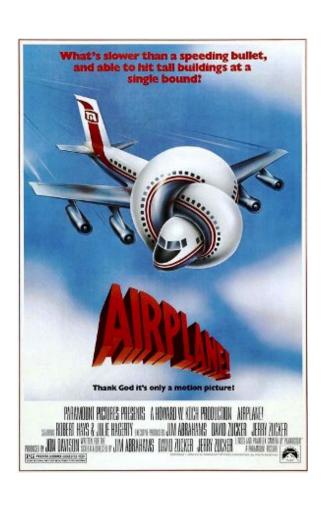# US Airlines Costs

Project 1: Multiple Linear Regression

Presented by Lacroix Scott, GADS-NYC12

MLR Problem Description

- Question: Can we predict Airline costs by factors such as customer revenue, fuel price and time?

- Responsive variable:
  - Cost ('cost')
- Explanatory variables:
  - Customer revenue ('output')
  - Fuel Price ('fuel')
  - Year ('year ')

Source: http://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Airline.html

Methodology

- Descriptive Statistics

variables	Y	X1	X2	X3
description	cost	output	fuel	year
count	90	90.0	90	90.0
mean	1,122,524	0.5	471,683	8.0
std	1,192,075	0.5	329,503	4.3
min	68,978	0.0	103,795	1.0
25%	292,046	0.1	129,848	4.0
50%	637,001	0.3	357,434	8.0
75%	1,345,968	0.9	849,840	12.0
max	4,748,320	1.9	1,015,610	15.0

- Created different training sets from test_train_split in Sklearn.cross_validation
- Fit the training data

Results

- Regressor Score: 91%