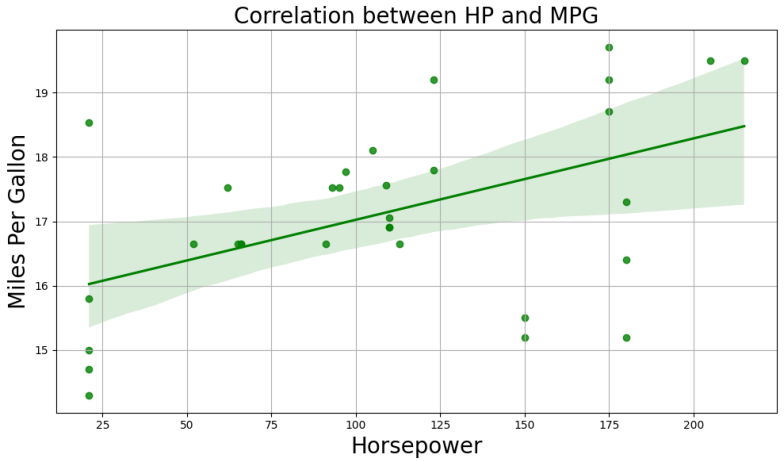
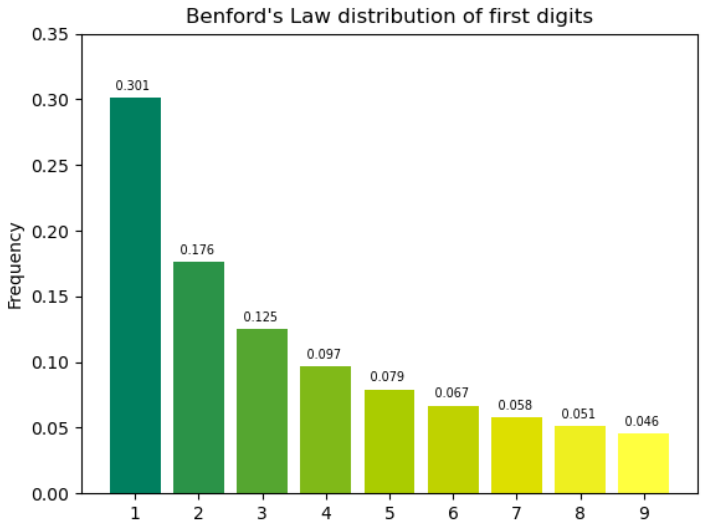
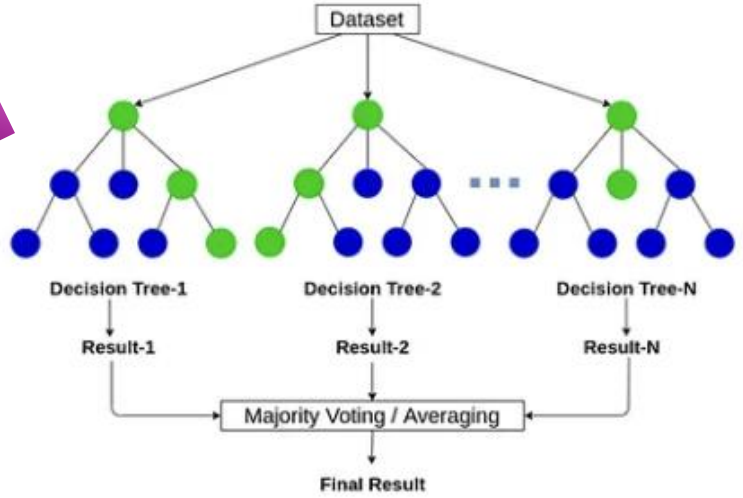
Detection of Correlation Reversal Manipulation via Benford's Law and Random Forest





Benford's Law
(Statistical Analysis)

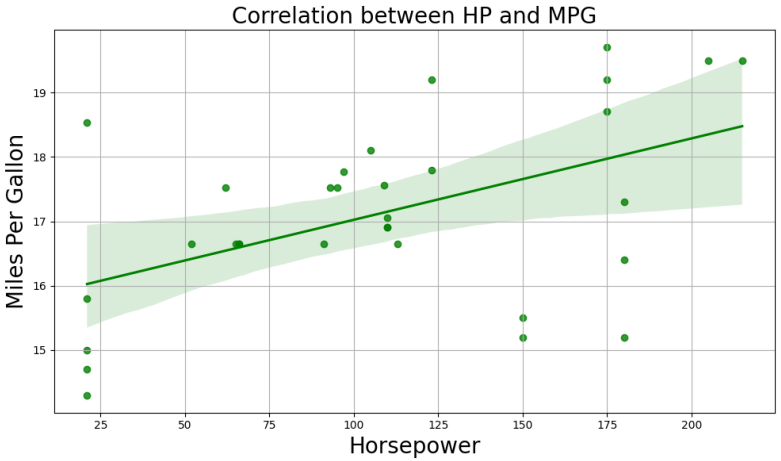


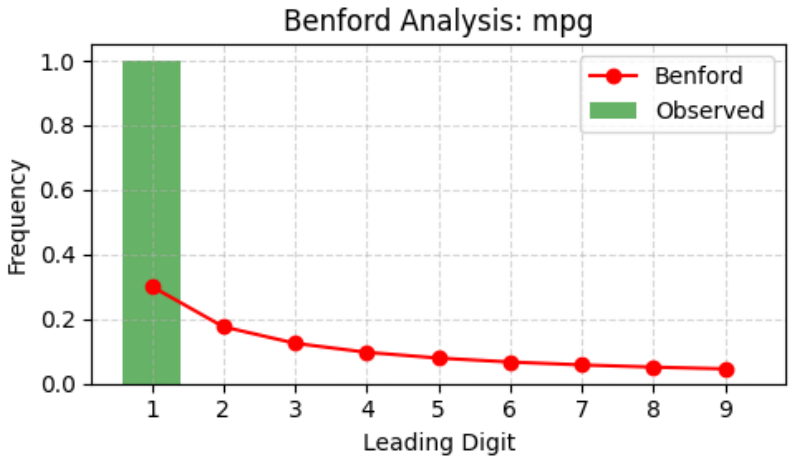
Random Forest
(Machine Learning)

Detect

1. Data Manipulation
2. Correlation Reversal

Detect Manipulation using Benford's Law





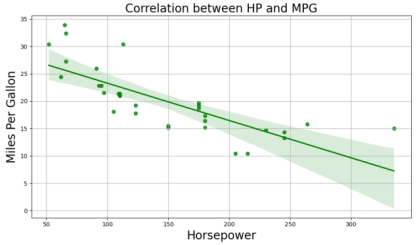
Chi-square = 74.3017, p-value = 0.0000000000006

Criteria: Chi-square > 30.58 and p < 0.0001

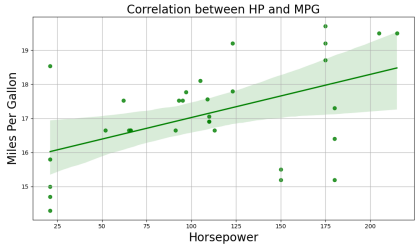
Classified as Banford Anomaly and extreme

Detect Manipulation using Random Forest

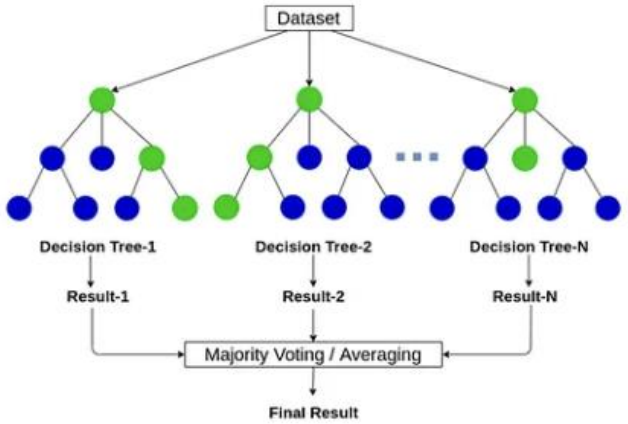
Original Data (Label=0)



Manipulated Data (Label=1)

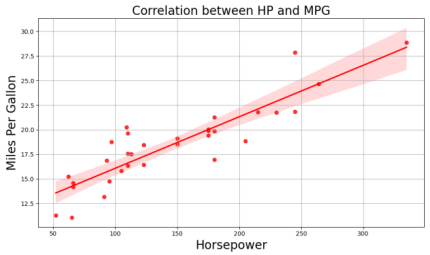


Supervised Training



RF Model

New Manipulated Data

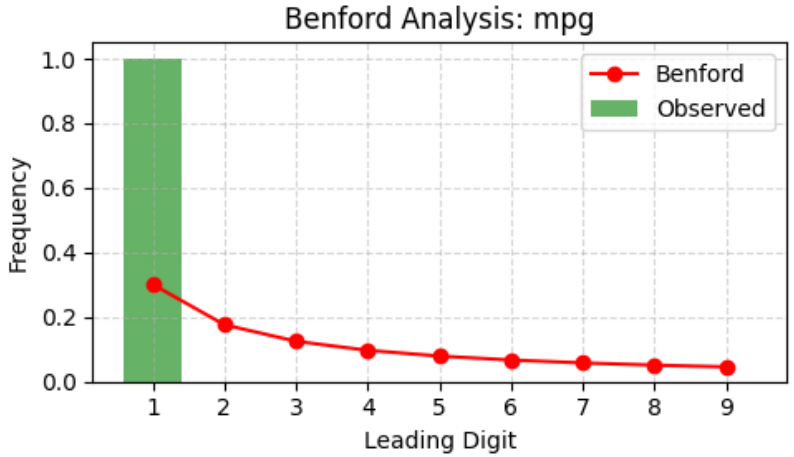


Classification

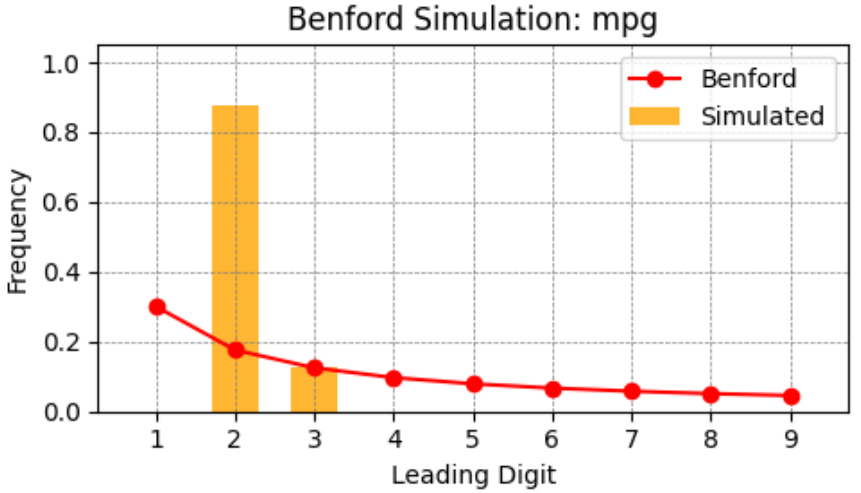
RF Model

As **manipulated**

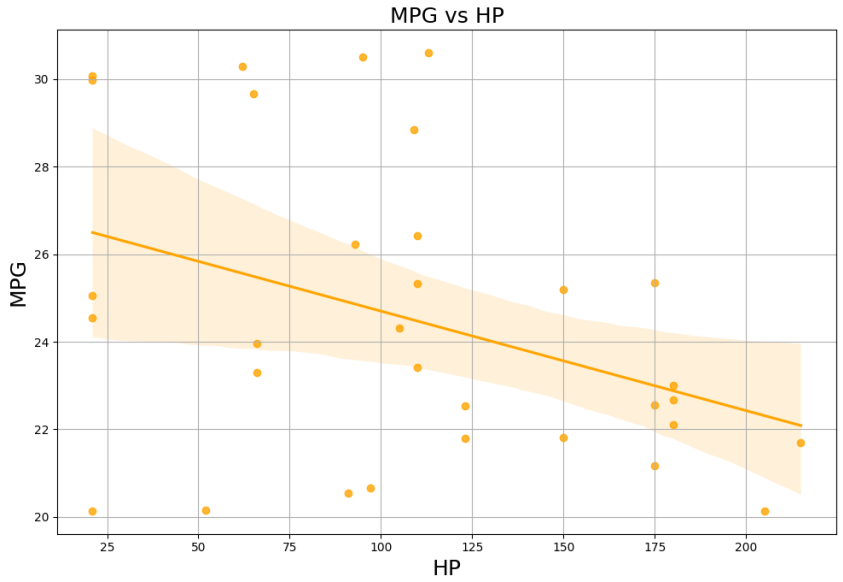
Detect Correlation Reversal using Benford's Law

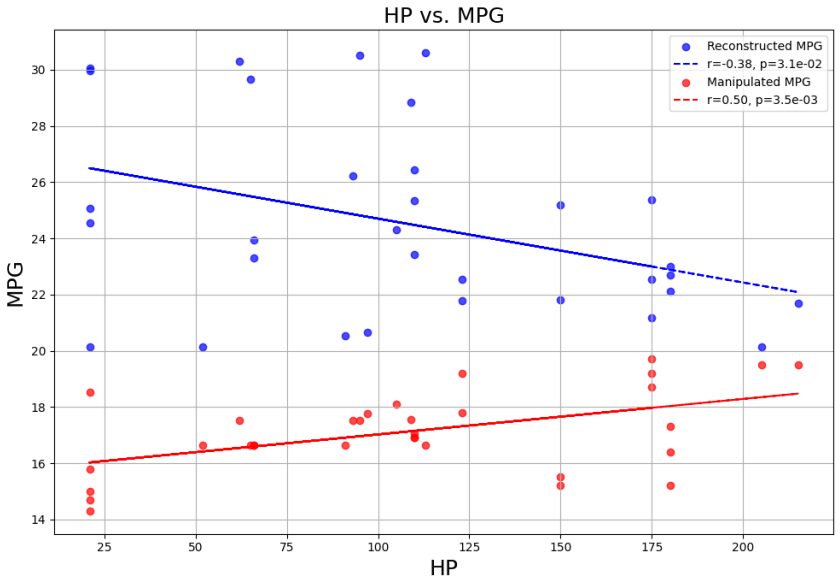


Stochastic Simulation



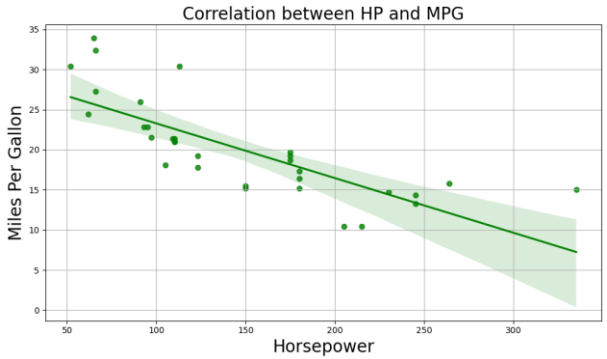
Reconstructed Data



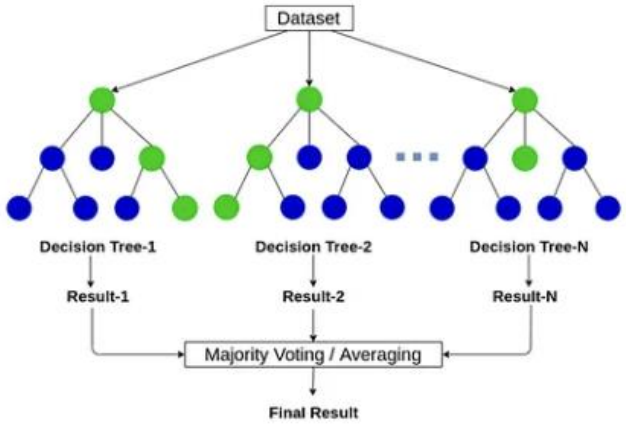


Detect Correlation Reversal using Random Forest

Original Data

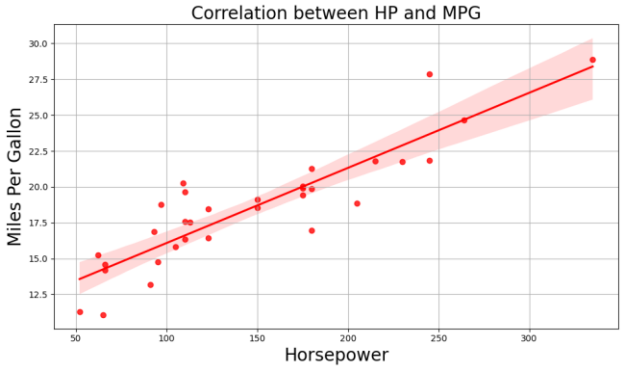


Training



RF Model

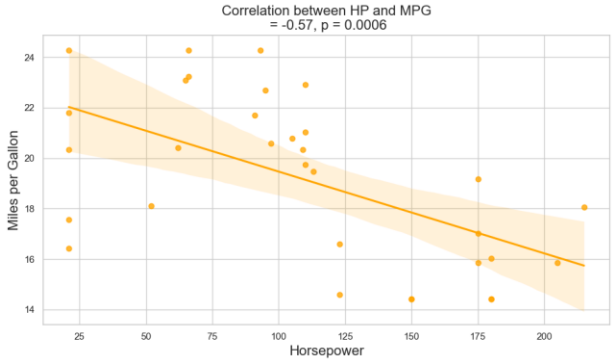
Manipulated Data

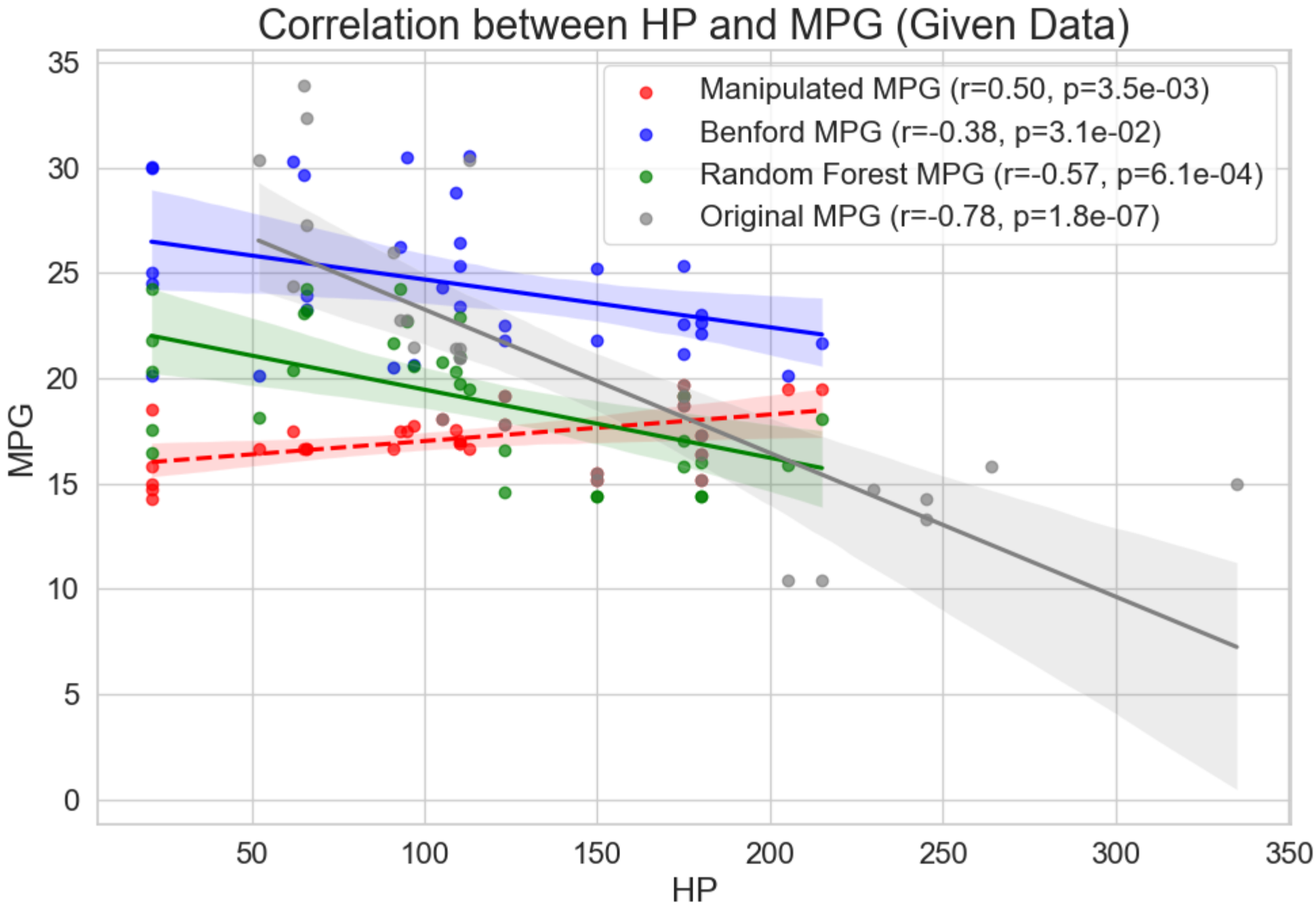


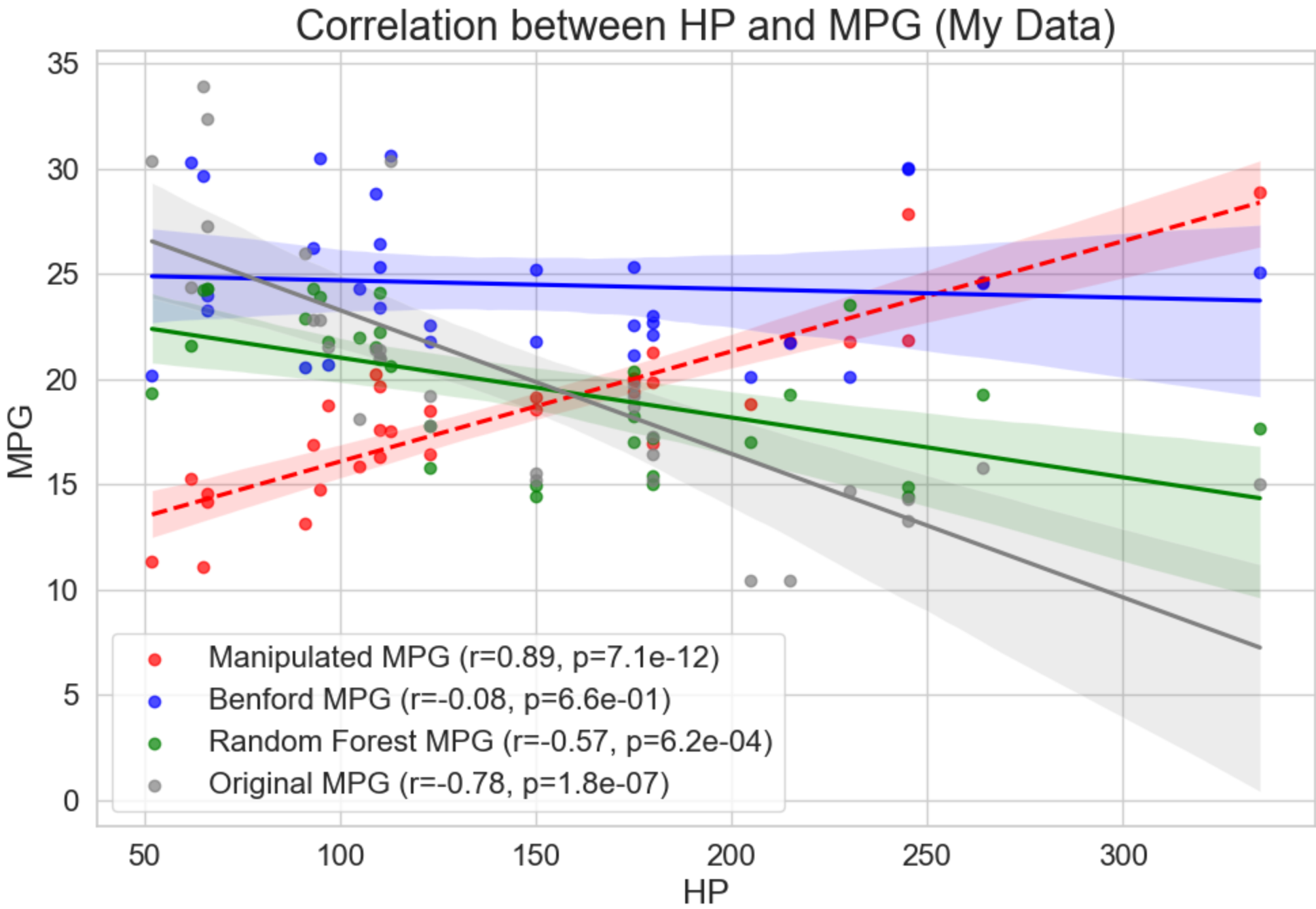
Regression

RF Model

Reconstructed Data




Correlation between HP and MPG (Given Data)
Manipulated MPG (r=0.50, p=3.5e-03)
Benford MPG (r=-0.38, p=3.1e-02)
Random Forest MPG (r=-0.57, p=6.1e-04)
Original MPG (r=-0.78, p=1.8e-07)
MPG
HP
0
5
10
15
20
25
30
35
50
100
150
200
250
300
350


Correlation between HP and MPG (My Data)
MPG
HP
0
5
10
15
20
25
30
35
50
100
150
200
250
300
Manipulated MPG (r=0.89, p=7.1e-12)
Benford MPG (r=-0.08, p=6.6e-01)
Random Forest MPG (r=-0.57, p=6.2e-04)
Original MPG (r=-0.78, p=1.8e-07)