

# LINEAR REGRESSION

## LINEAR REGRESSION

### TEAM MEMBERS:

- SARITA JOSHI and SHARMODEEP SARKAR

### IMPLEMENTATION :

---

This code is implemented using Map-reduce infrastructure for evaluating the linear regression for the average price of the Airlines w.r.t. to variables

- AirTimea
- Distance

For almost all the airlines, the root mean square error for the Air-Time vs Average-Ticket-Price graph is less than the root mean square error for the Distance-Traveled vs Average-Ticket-Price graph. This proves that Air-Time is a better predictor variable than Distance Traveled for the Criterion variable (Average Ticket Price).

Calculating the average ticket price from the slopes and the y-intercepts of the Air-Time regression graphs of various Airlines, we conclude F9 has the least operating cost. The sequence of Airlines (cheapest to dearest) is below : Order is as below:

Y	M (slope)	X	C (y-intercept)	CARRIER
161.5555004	0.31315308	200	98.92488449	F9
253.6347803	1.169361666	200	19.76244717	AS

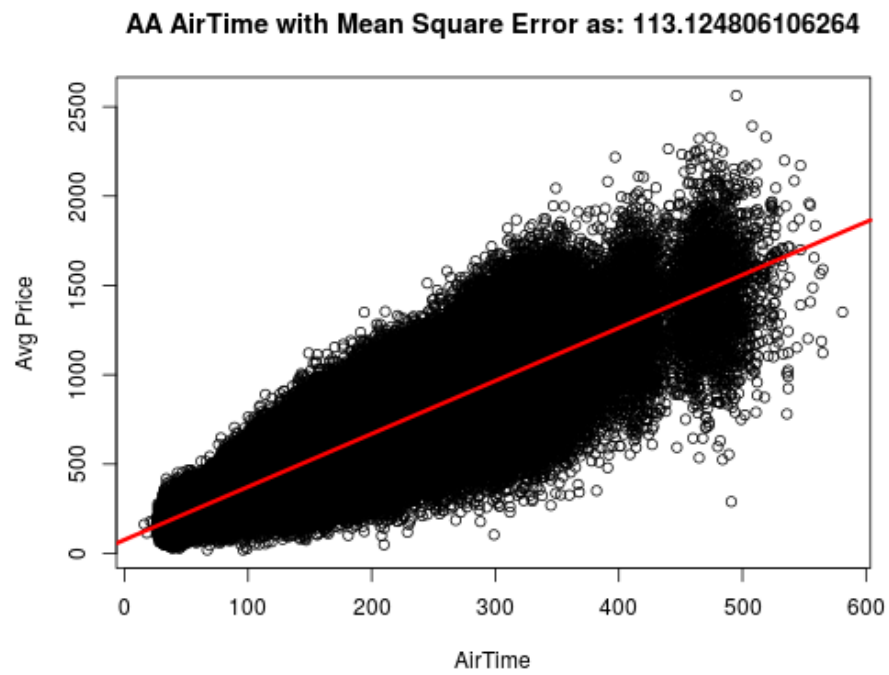
254.1749324 1.151380719 200 23.89878863 WN  
643.7535826 3.030234001 200 37.70678241 HA  
651.7718424 2.907244838 200 70.32287471 MQ  
657.5720876 2.907895626 200 75.99296241 EV  
658.3576378 2.944299082 200 69.49782139 OO  
664.2111451 2.950446718 200 74.12180139 B6  
664.4685275 3.002792134 200 63.91010068 VX  
669.3963743 2.964938324 200 76.40870959 AA  
870.03825 3.84003498 200 102.0312539 US  
885.8092366 3.895228718 200 106.7634929 DL  
1096.941487 4.76342438 200 144.2566111 UA

\*\*\* STEPS TO BE EXECUTED in below SEQUENCE FOR AUTOMATED  
GENERATOON OF myReport.pdf

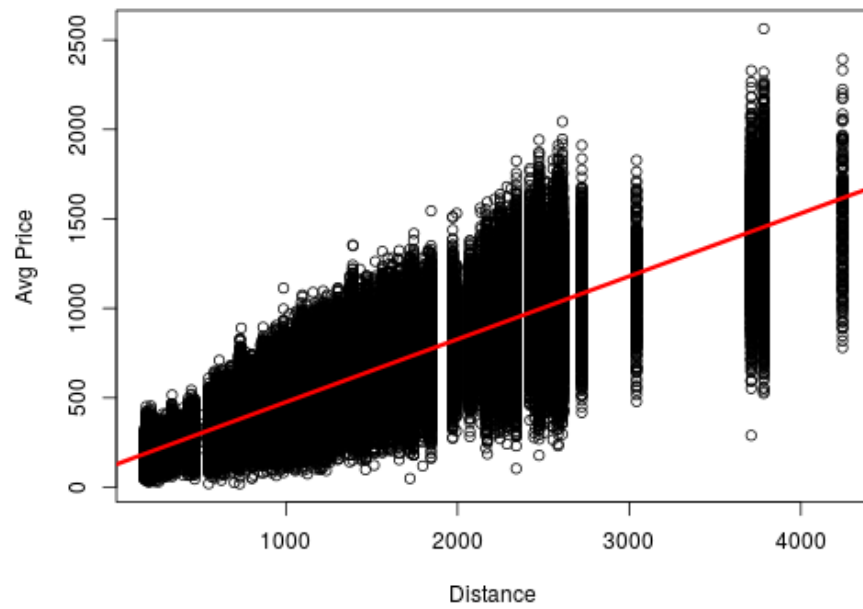
- make jar
- make permissiona
- make hstart
- make pseudo
- make hstop
- make clean
- make emr
- make report

## GRAPHICAL REPRESENTATION

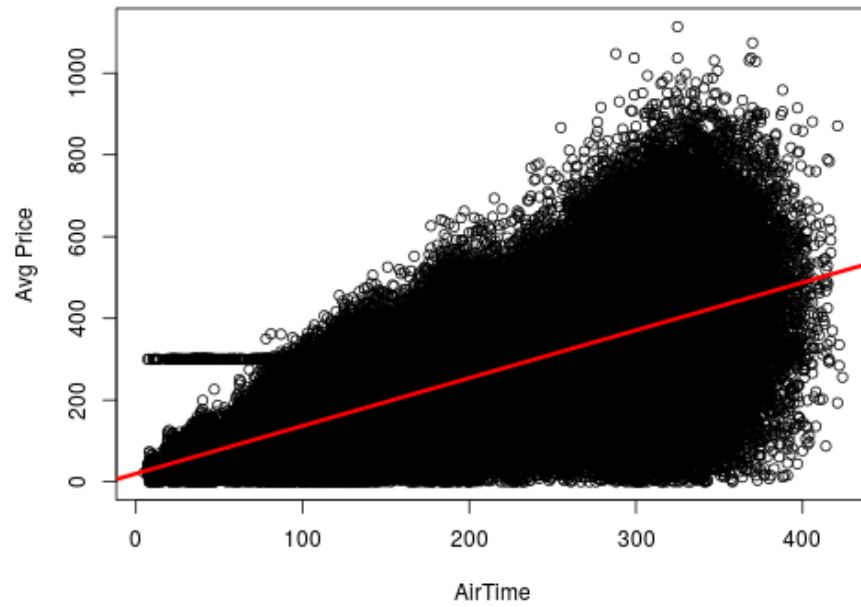
LINEAR REGRESSION -> Average Price with distance and Average Price with Air time:



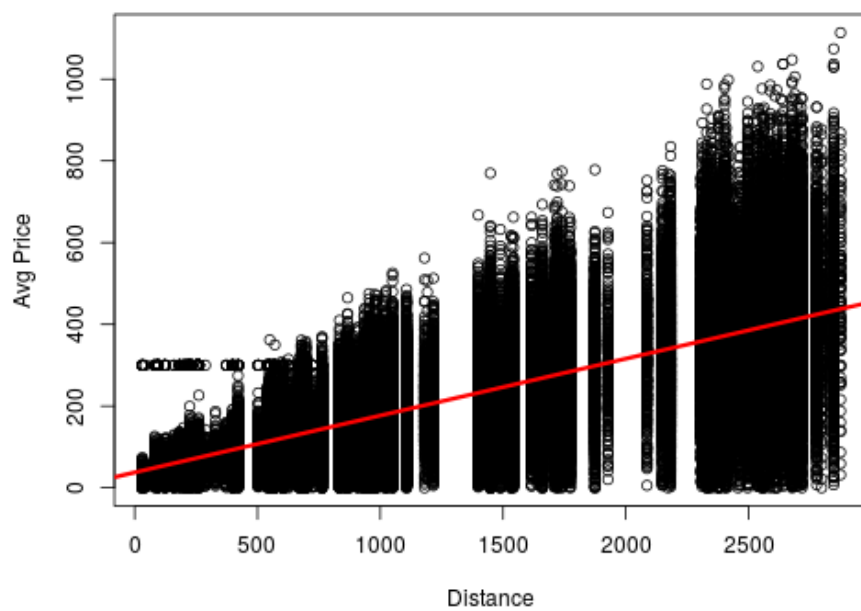
**AA Distance with Mean Square Error as: 116.644118615511**



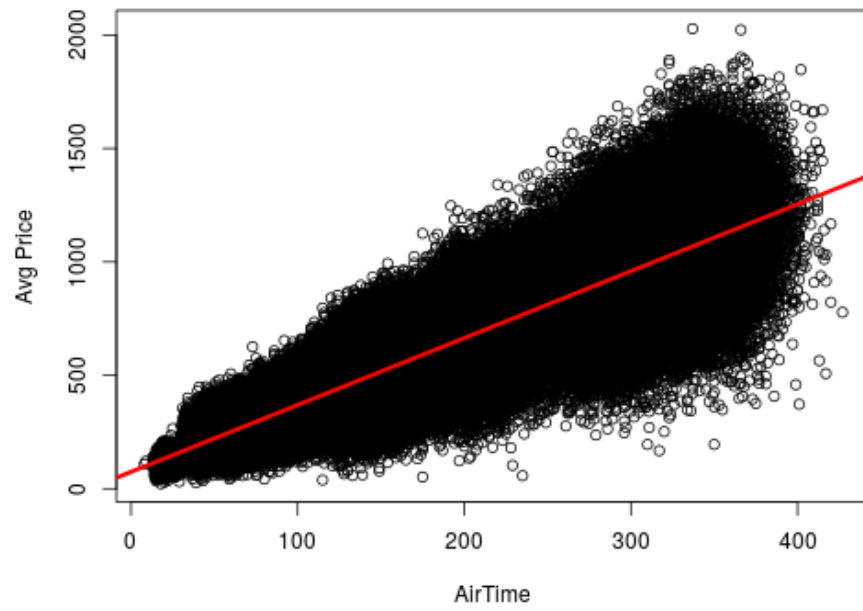
**AS AirTime with Mean Square Error as: 92.4223309701224**



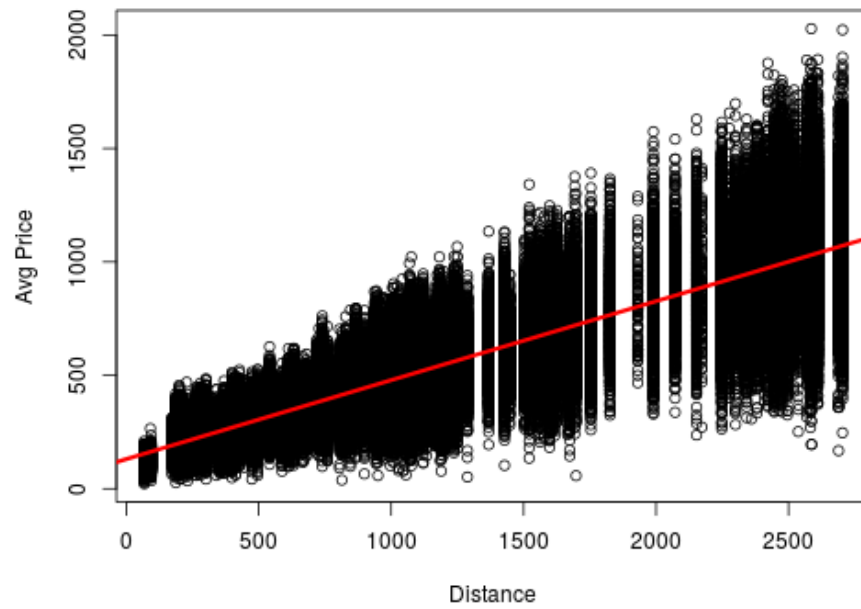
AS Distance with Mean Square Error as: 92.7256704725164



**B6 AirTime with Mean Square Error as: 115.274269602097**

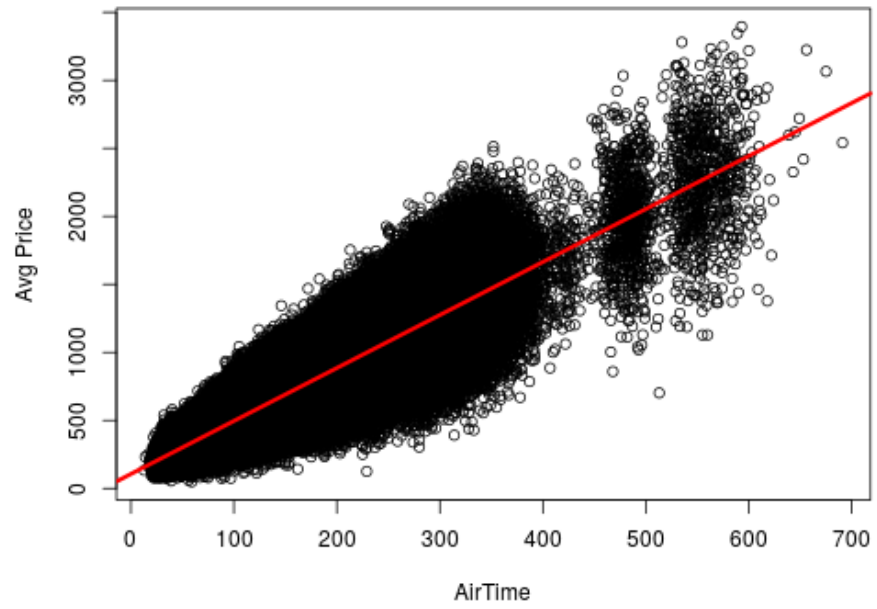


**B6 Distance with Mean Square Error as: 118.979827342687**

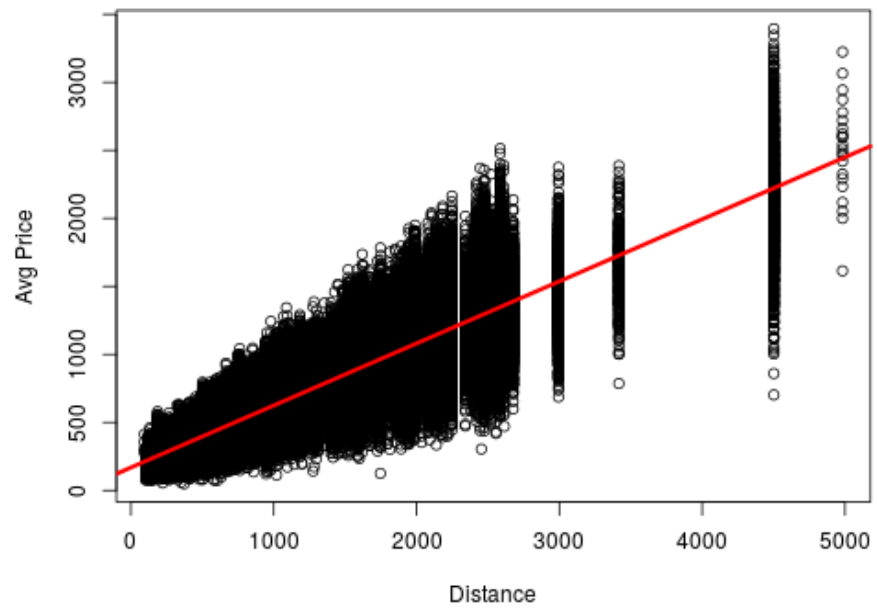




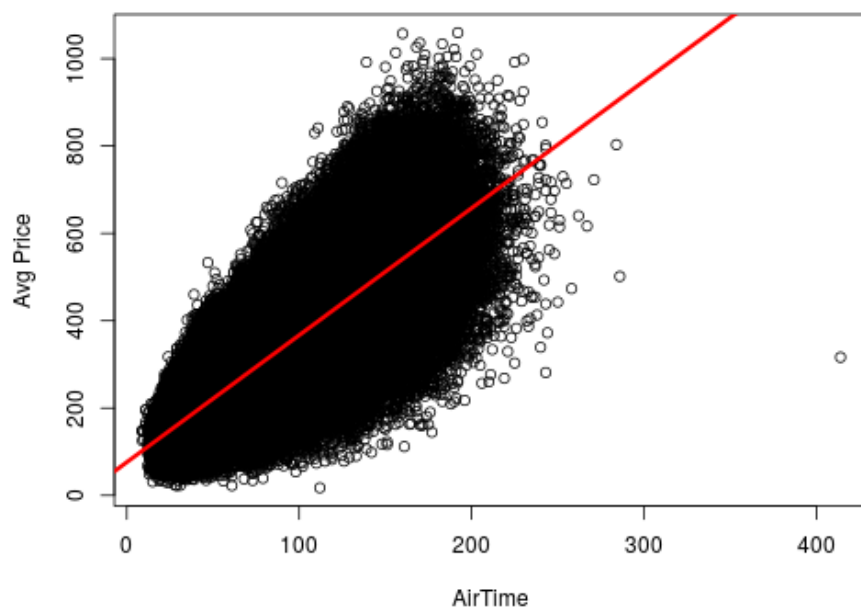
**DL AirTime with Mean Square Error as: 117.051790558421**



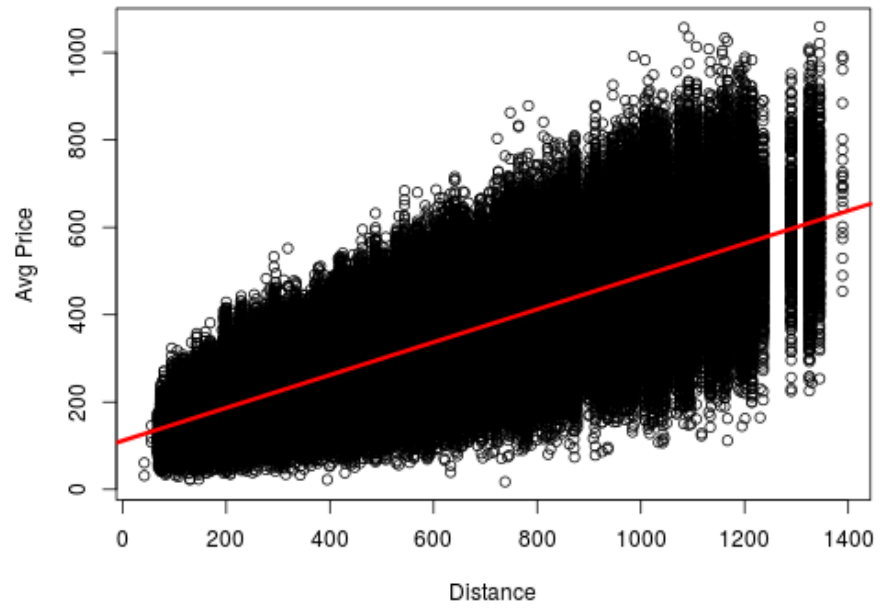
DL Distance with Mean Square Error as: 121.424101773891



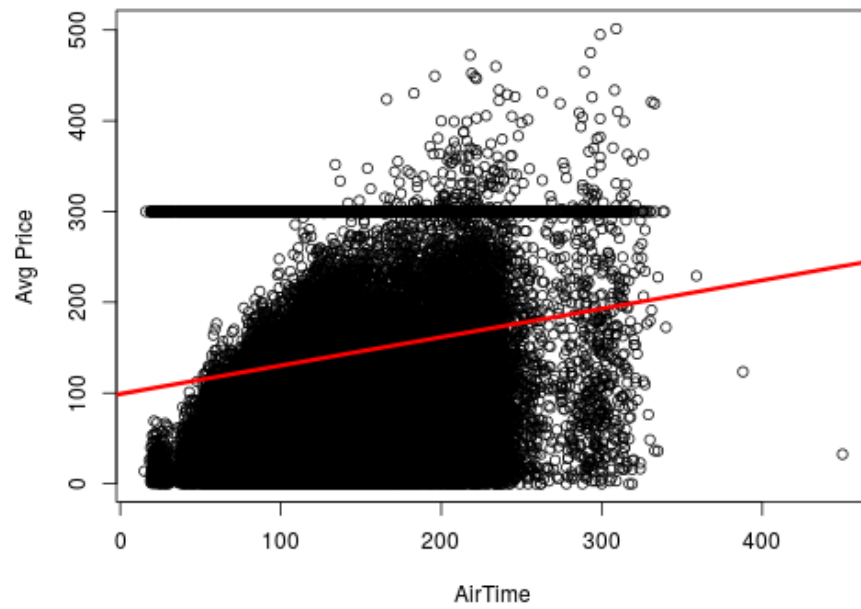
**EV AirTime with Mean Square Error as: 66.3278350361565**



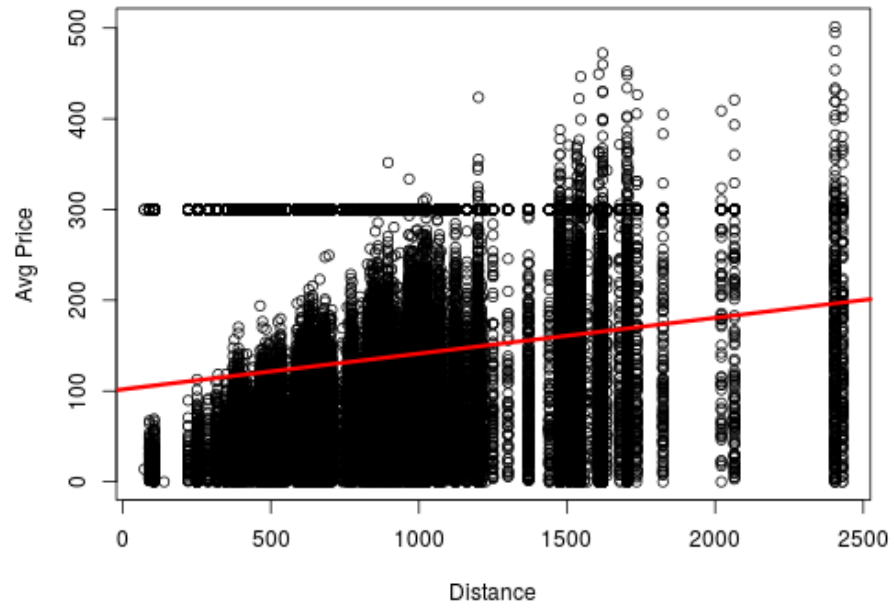
EV Distance with Mean Square Error as: 66.5861152464683



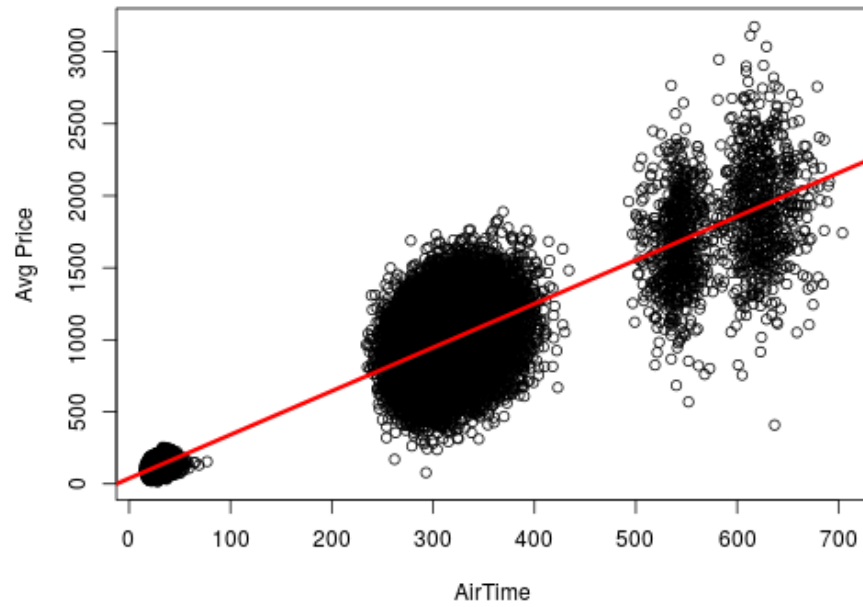
**F9 AirTime with Mean Square Error as: 116.641094636089**



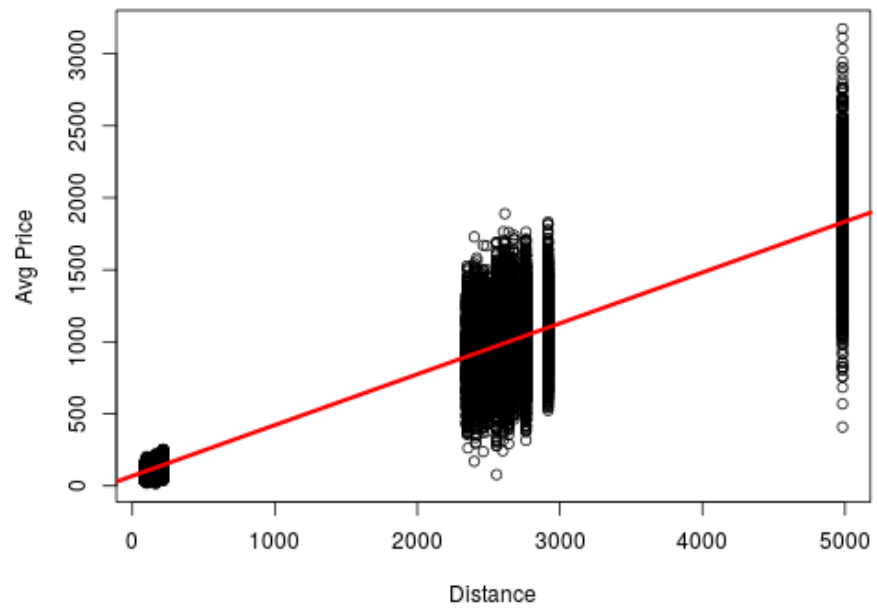
F9 Distance with Mean Square Error as: 116.644650239856



**HA AirTime with Mean Square Error as: 92.4685186778084**

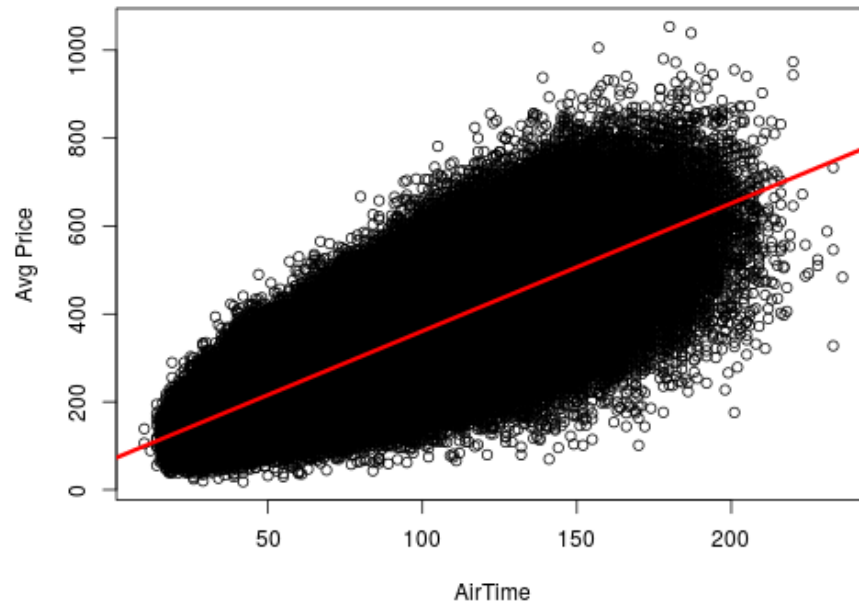


HA Distance with Mean Square Error as: 93.0387899294544

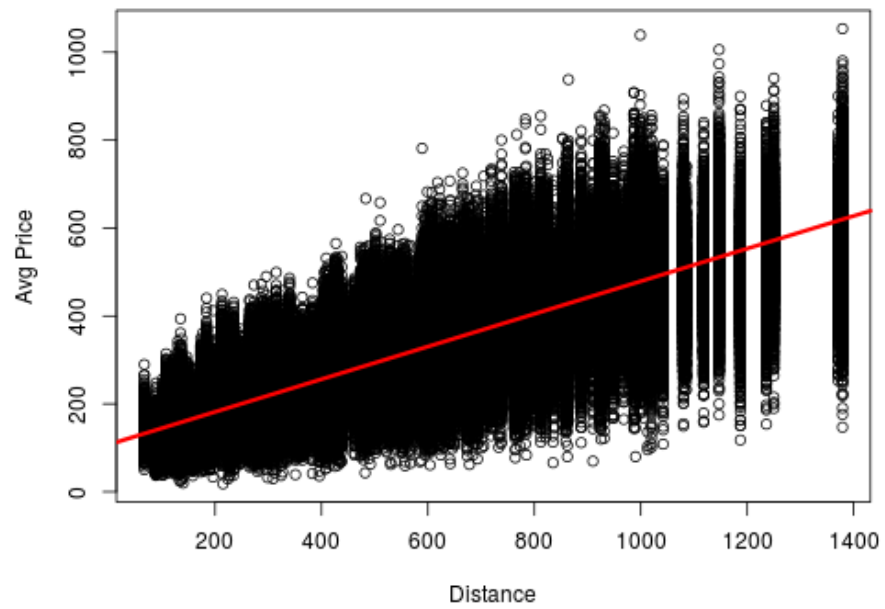




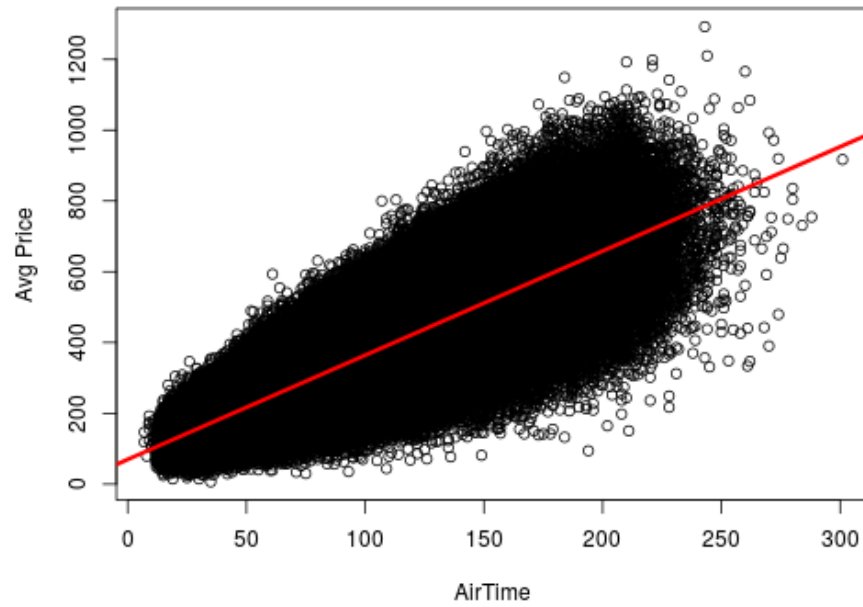
**MQ AirTime with Mean Square Error as: 65.5630178530965**



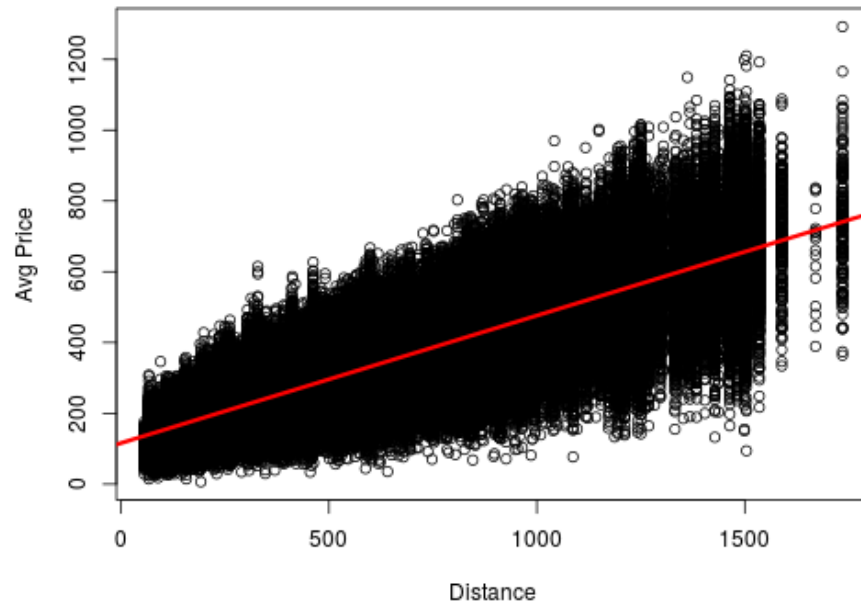
**MQ Distance with Mean Square Error as: 66.9017166260133**



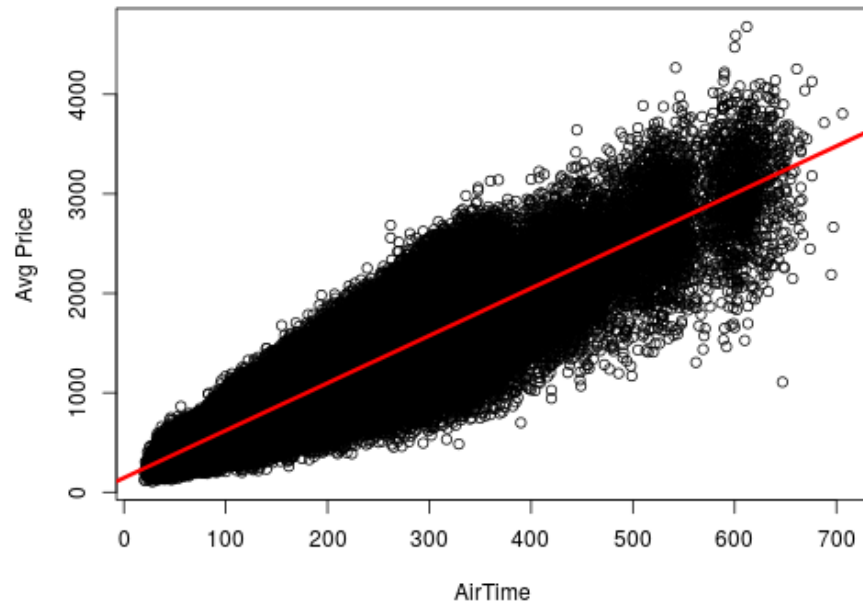
**OO AirTime with Mean Square Error as: 65.2126058026948**



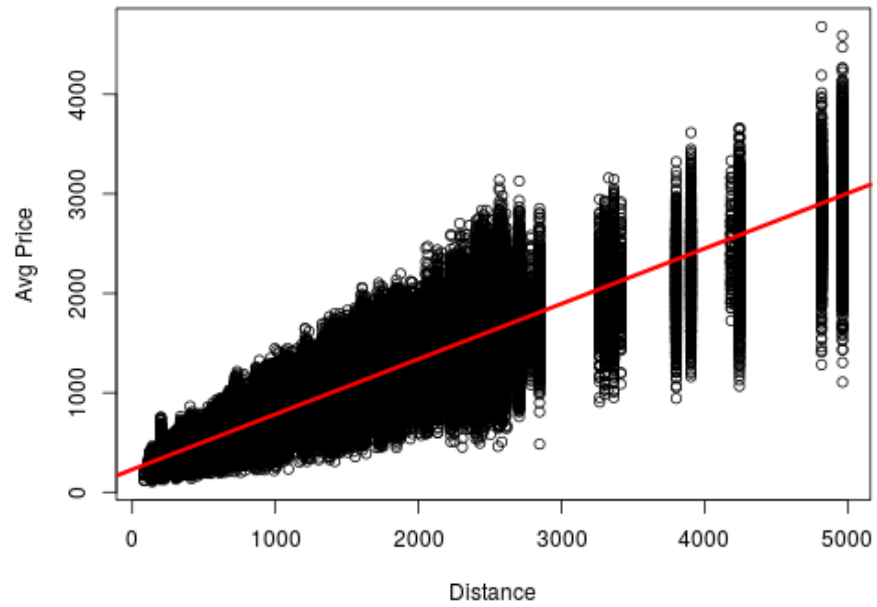
OO Distance with Mean Square Error as: 65.6559313649689



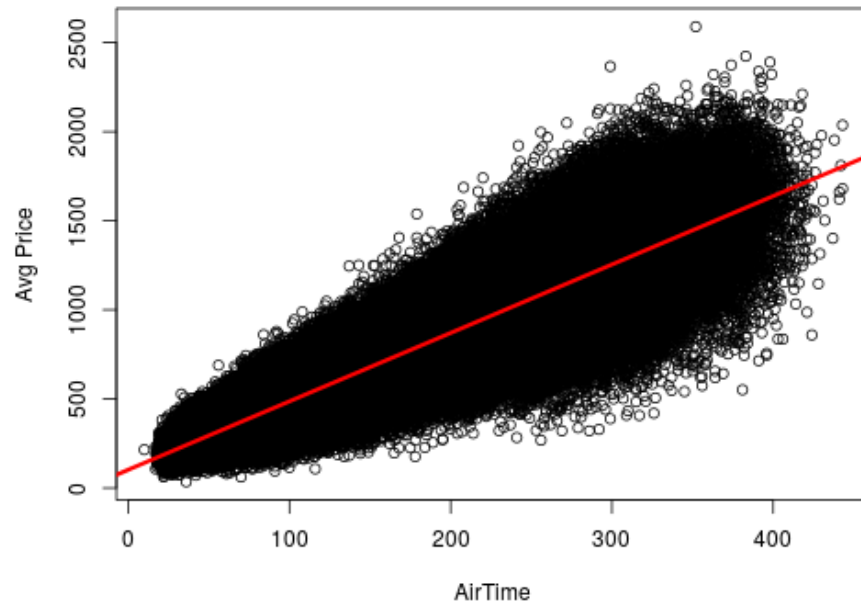
**UA AirTime with Mean Square Error as: 178.120172627656**



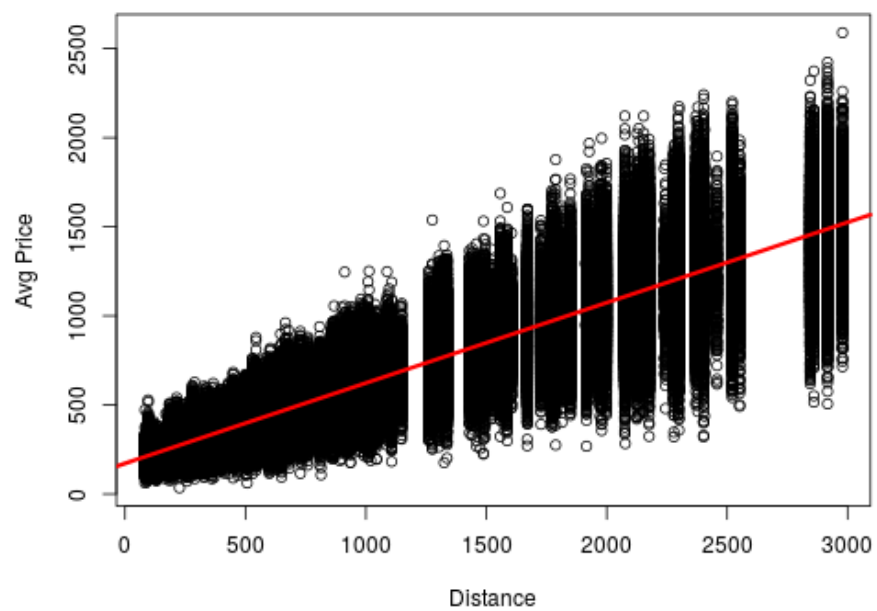
UA Distance with Mean Square Error as: 186.528677506783



**US AirTime with Mean Square Error as: 119.700325282593**

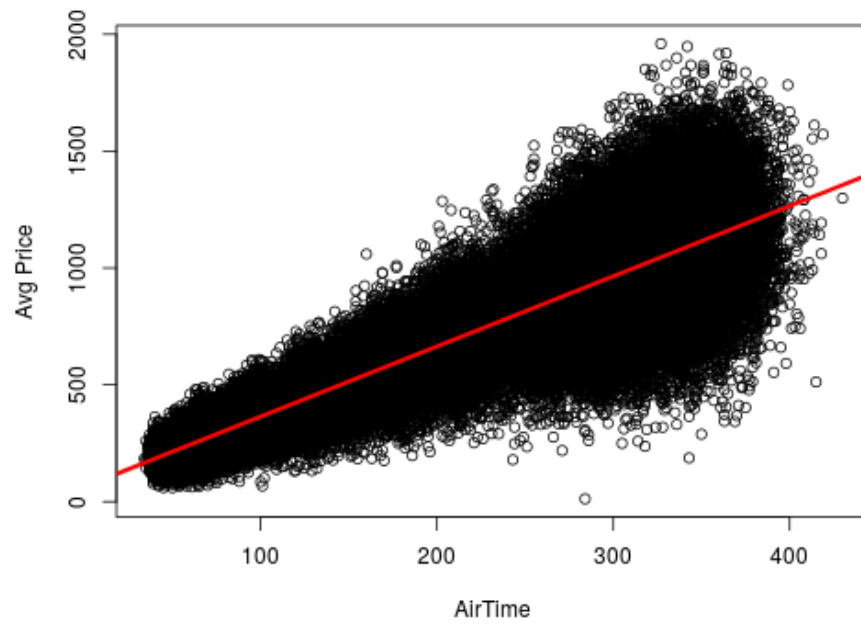


US Distance with Mean Square Error as: 123.694077580289

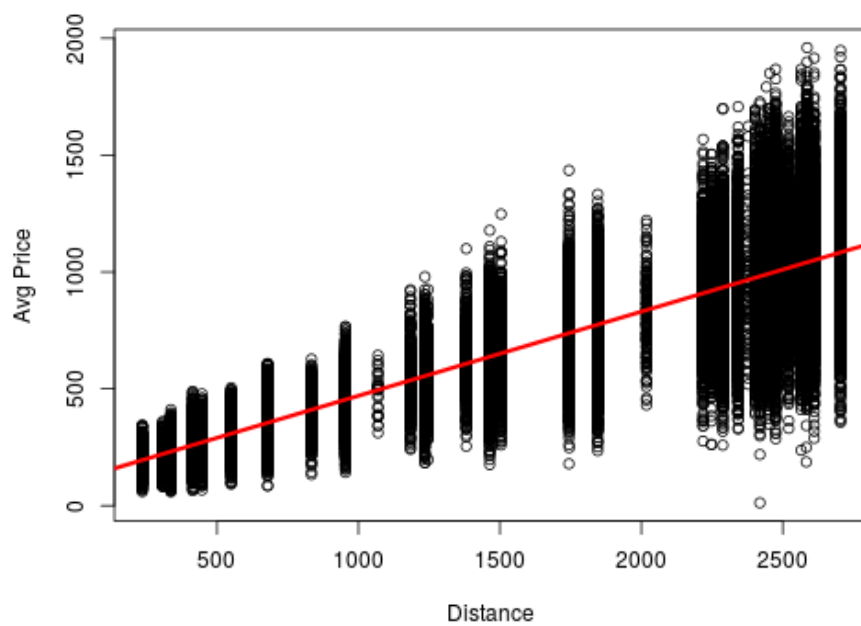




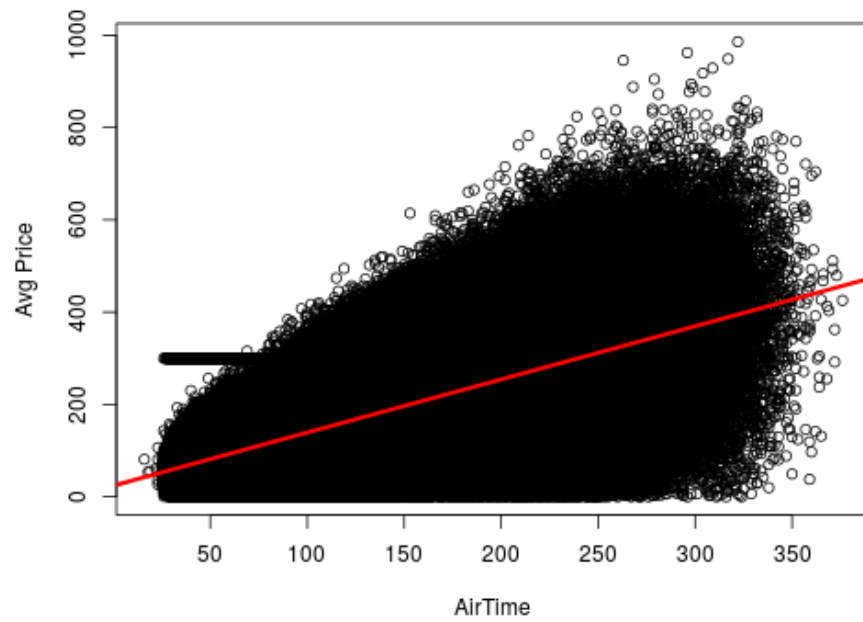
**VX AirTime with Mean Square Error as: 147.669737865079**

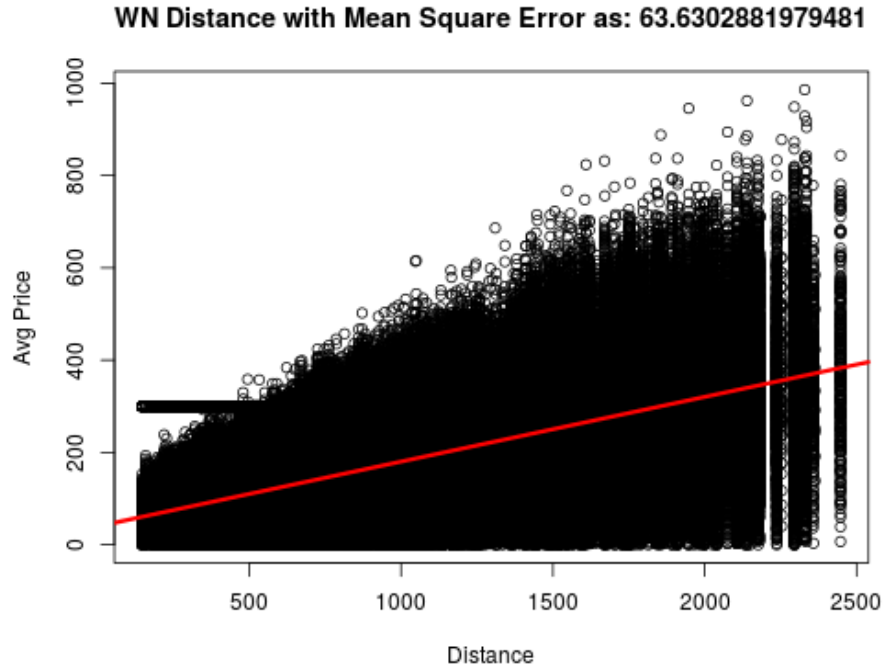


VX Distance with Mean Square Error as: 154.107589050741



**WN AirTime with Mean Square Error as: 63.2500731693329**





## Mathematical Analysis

- [1] "2.96493832370479 is the slope for time graph for the Airline AA"
- [2] "1.1693616655077 is the slope for time graph for the Airline AS"
- [3] "2.95044671838792 is the slope for time graph for the Airline B6"
- [4] "3.89522871839555 is the slope for time graph for the Airline DL"
- [5] "2.90789562595067 is the slope for time graph for the Airline EV"
- [6] "0.313153079599492 is the slope for time graph for the Airline F9"
- [7] "3.03023400104918 is the slope for time graph for the Airline HA"
- [8] "2.90724483836948 is the slope for time graph for the Airline MQ"
- [9] "2.94429908188421 is the slope for time graph for the Airline OO"
- [10] "4.76342437964737 is the slope for time graph for the Airline UA"
- [11] "3.84003498039684 is the slope for time graph for the Airline US"
- [12] "3.00279213400312 is the slope for time graph for the Airline VX"
- [13] "1.15138071871333 is the slope for time graph for the Airline WN"
- [1] "76.408709585957 is the y-intercept for time graph for the Airline AA"
- [2] "19.7624471679235 is the y-intercept for time graph for the Airline AS"
- [3] "74.1218013864907 is the y-intercept for time graph for the Airline B6"
- [4] "106.763492896288 is the y-intercept for time graph for the Airline DL"
- [5] "75.9929624074941 is the y-intercept for time graph for the Airline EV"

- [6] “98.9248844866689 is the y-intercept for time graph for the Airline F9”
- [7] “37.7067824121906 is the y-intercept for time graph for the Airline HA”
- [8] “70.3228747134723 is the y-intercept for time graph for the Airline MQ”
- [9] “69.4978213895832 is the y-intercept for time graph for the Airline OO”
- [10] “144.256611084893 is the y-intercept for time graph for the Airline UA”
- [11] “102.031253877006 is the y-intercept for time graph for the Airline US”
- [12] “63.9101006807591 is the y-intercept for time graph for the Airline VX”
- [13] “23.8987886278092 is the y-intercept for time graph for the Airline WN”