Sort Paper

Purpose

See: Sort paper

- To take performance data on the insertion sort, selection sort, heap sort, merge sort, and quick sort as a function of the number of values sorted.
- To do a least squares curve fit on the data with an n^2 model and an $n \lg n$ model.
- To state the theoretical Θ increase in execution time as a function of the number of values sorted.
- To report whether or not the data confirms the theory.
- To recommend the best of the five sort algorithms analyzed.
- To answer any other interesting questions about the data that may arise from your analysis.

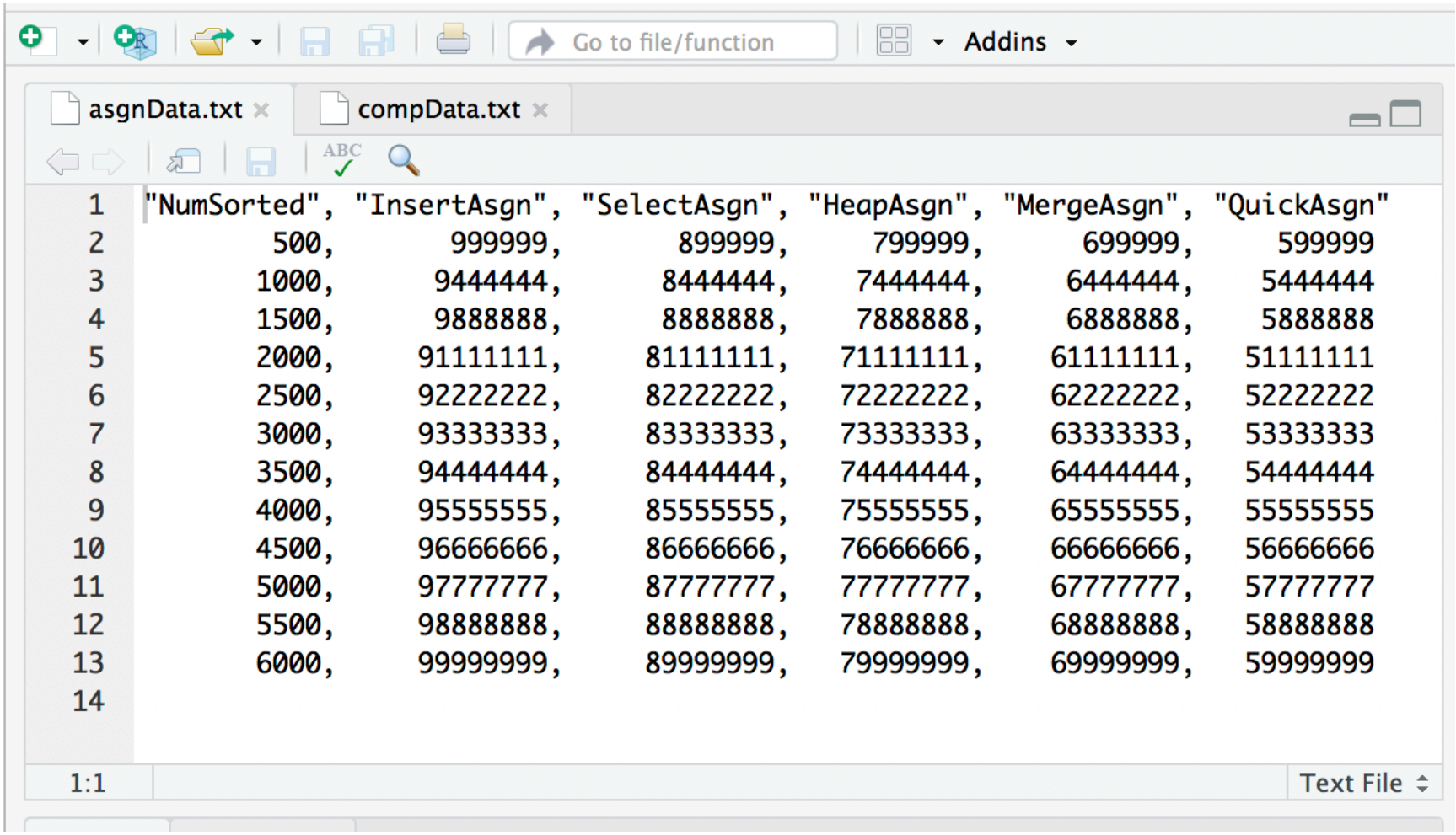
Requirement

Analyze your data using R with RStudio.

See:

- Setup for RStudio
- Data management in RStudio
- Plotting raw data in RStudio
- Curve fitting in RStudio

Data management in RStudio



	"NumSorted"	"InsertAsgn"	"SelectAsgn"	"HeapAsgn"	"MergeAsgn"	"QuickAsgn"
2	500,	999999,	899999,	799999,	699999,	599999
3	1000,	9444444,	8444444,	7444444,	6444444,	5444444
4	1500,	9888888,	8888888,	7888888,	6888888,	5888888
5	2000,	91111111,	81111111,	71111111,	61111111,	51111111
6	2500,	92222222,	82222222,	72222222,	62222222,	52222222
7	3000,	93333333,	83333333,	73333333,	63333333,	53333333
8	3500,	94444444,	84444444,	74444444,	64444444,	54444444
9	4000,	95555555,	85555555,	75555555,	65555555,	55555555
10	4500,	96666666,	86666666,	76666666,	66666666,	56666666
11	5000,	97777777,	87777777,	77777777,	67777777,	57777777
12	5500,	98888888,	88888888,	78888888,	68888888,	58888888
13	6000,	99999999,	89999999,	79999999,	69999999,	59999999
14						

Curve fitting in RStudio

Question

Is a plot of a performance metric vs. n

$$T(n) = \Theta(n^2)$$

or is it

$$T(n) = \Theta(n \lg n)$$

Curve fitting in RStudio

Answer

Calculate the minimum residual standard error (RSE) for each possibility

$$RSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{d.f.}}$$

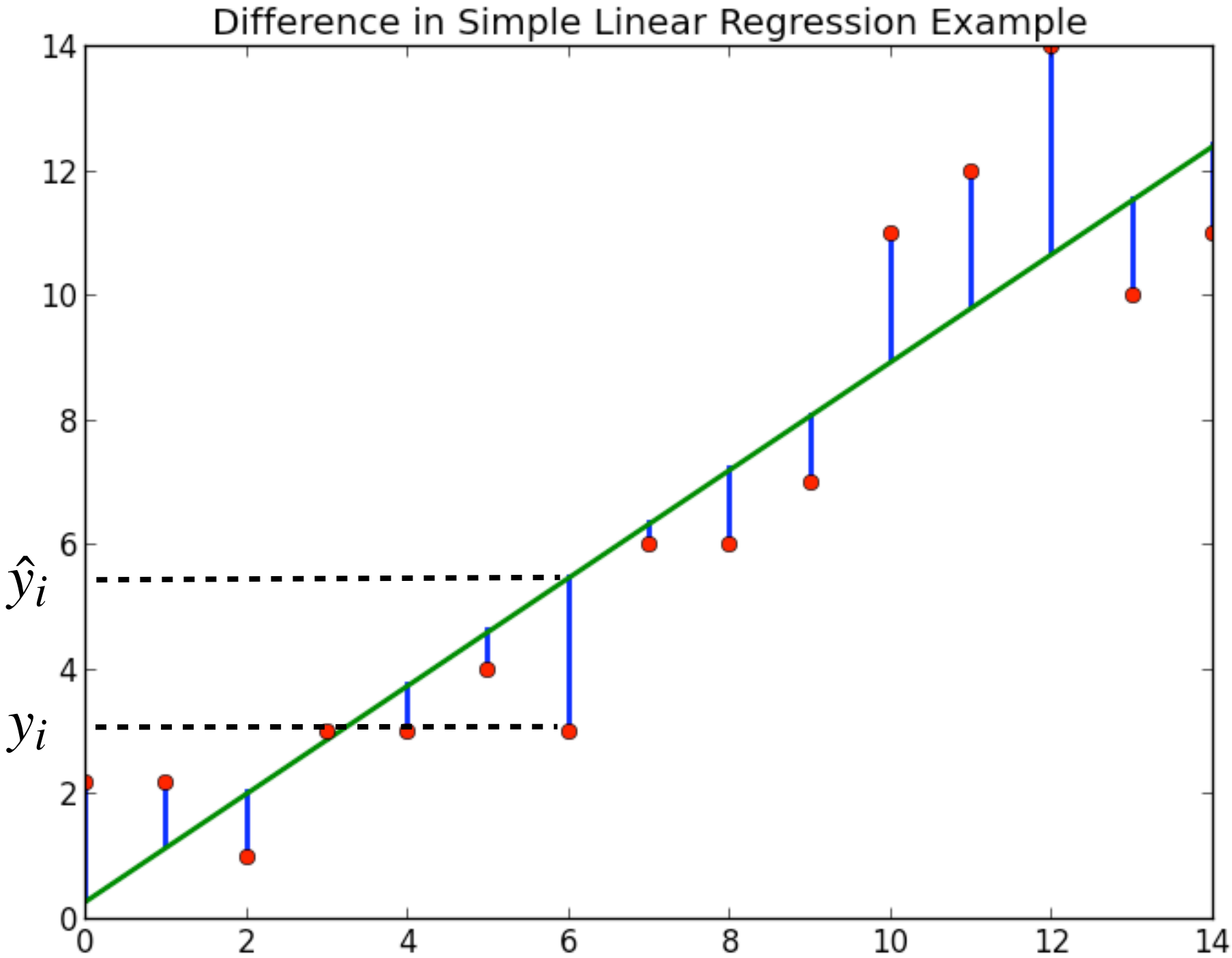
Curve fitting in RStudio

Example of a linear fit:

The equation for a line is

$$y = Ax + B$$

So, adjust A and B to minimize the RSE.


Difference in Simple Linear Regression Example
$\hat{y}_i$
y_i
0
2
4
6
8
10
12
14

Curve fitting in RStudio

$$RSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{d.f.}}$$

d.f. is degrees of freedom
= (number of data points) – (number of coefficients)

Curve fitting in RStudio

For a line, adjust A and B in

$$y = Ax + B$$

to minimize the RSE.

Curve fitting in RStudio

Compute RSE for two possibilities:

$$y = An^2 + Bn + C$$

$$y = An\lg n + Bn + C$$

Curve fitting in RStudio

Example: Quadratic fit

```
Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                          3416871       8122  420.71  < 2e-16 ***
poly(dataFrame[, indepVarName], 2)1  9817068      28134  348.93  < 2e-16 ***
poly(dataFrame[, indepVarName], 2)2  2324549      28134   82.62 2.82e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28130 on 9 degrees of freedom
Multiple R-squared:  0.9999,Adjusted R-squared:  0.9999
F-statistic: 6.429e+04 on 2 and 9 DF,  p-value: < 2.2e-16
```

Curve fitting in RStudio

Example: *n* lg *n* fit

```
Coefficients:
                                                    Estimate Std. Error t value
(Intercept)                                       1327296.56  241061.68   5.506
dataFrame[, indepVarName]                          -10547.80     831.28 -12.689
dataFrame[, indepVarName]:log(dataFrame[, indepVarName])    1357.47      92.53  14.671
                                                  Pr(>|t|)
(Intercept)                                       0.000377 ***
dataFrame[, indepVarName]                         4.78e-07 ***
dataFrame[, indepVarName]:log(dataFrame[, indepVarName]) 1.37e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 155300 on 9 degrees of freedom
Multiple R-squared:  0.9979,Adjusted R-squared:  0.9974
F-statistic:  2105 on 2 and 9 DF,  p-value: 9.571e-13
```

Curve fitting in RStudio

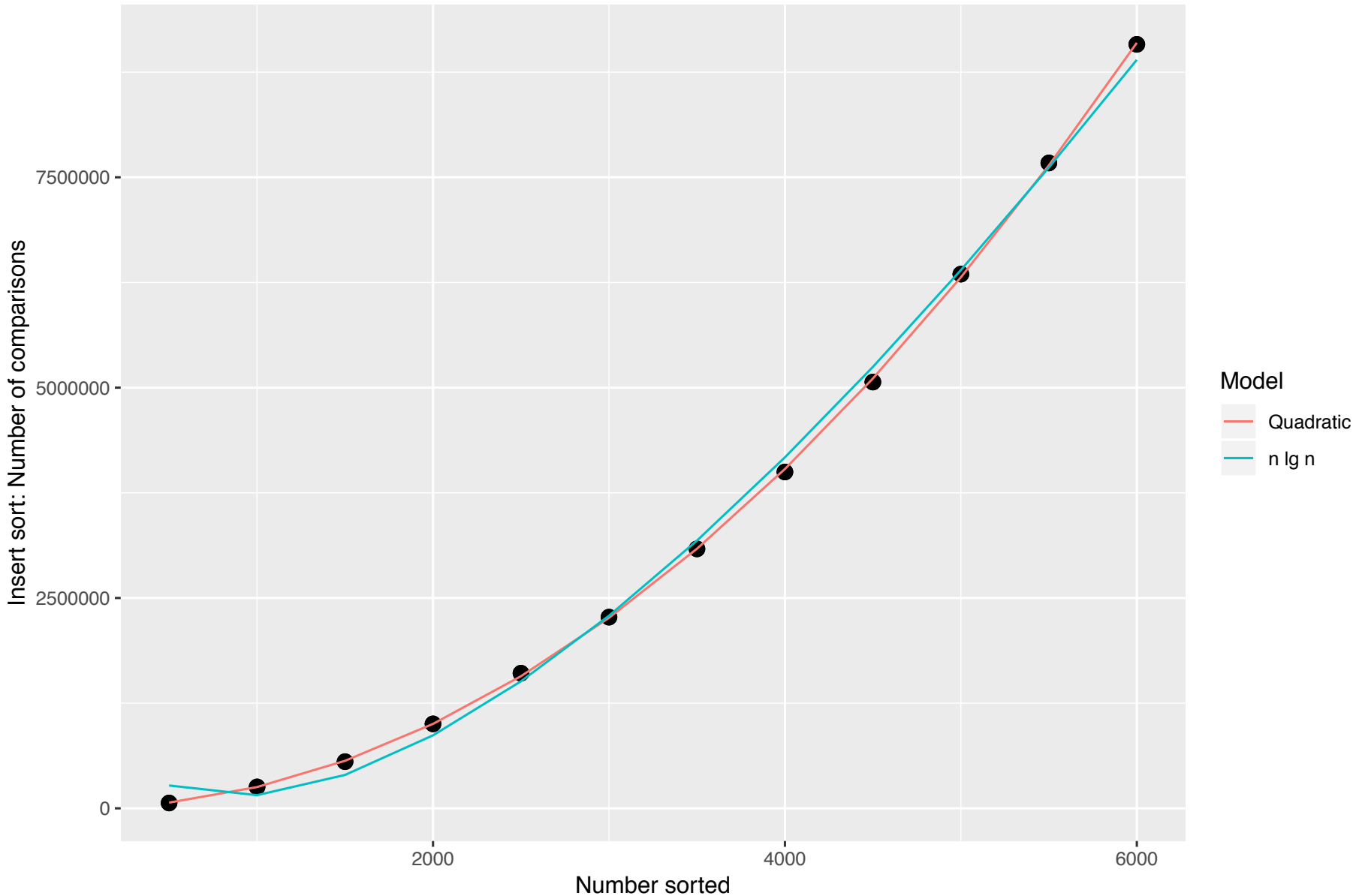
Compare

28,130 (quadratic) < 155,300 (n lg n)

Conclusion:

The performance is quadratic.

Curve fitting in RStudio



Requirement

Write your paper using LaTeX.

See:

- Setup for LaTeX
- The Not So Short Introduction to LaTeX2e

Demo LaTeX in Overleaf

IMPORTANT LaTeX Guideline

Click Typeset (Recompile) button OFTEN!

Design Patterns for Data Structures

Number of data points	Algorithm				
	Insert	Select	Heap	Merge	Quick
500					
1000					
1500					
2000					
2500					
3000					
3500					
4000					
4500					
5000					
5500					
6000					

Figure 1. Number of array element comparisons.

Design Patterns for Data Structures

Number of data points	Algorithm				
	Insert	Select	Heap	Merge	Quick
500					
1000					
1500					
2000					
2500					
3000					
3500					
4000					
4500					
5000					
5500					
6000					

Figure 2. Number of array element assignments.

Sort paper

Abstract

1. Introduction
2. Method
3. Results
4. Conclusion

References

Grading rubric
100 homework points

- 5 points: Form, LaTeX layout.
- 15 points: Grammar, punctuation, style.
- 5 points: Abstract.
- 5 points: Introduction.
- 20 points: Method.
- 40 points: Results.
- 5 points: Conclusion.
- 5 points: References.

See Sort paper for:

- Abstract and Conclusions
- Introduction
- Assertion - evidence
- Personal pronouns
- Present tense
- Active voice
- Conciseness
- Amount versus number
- Figures