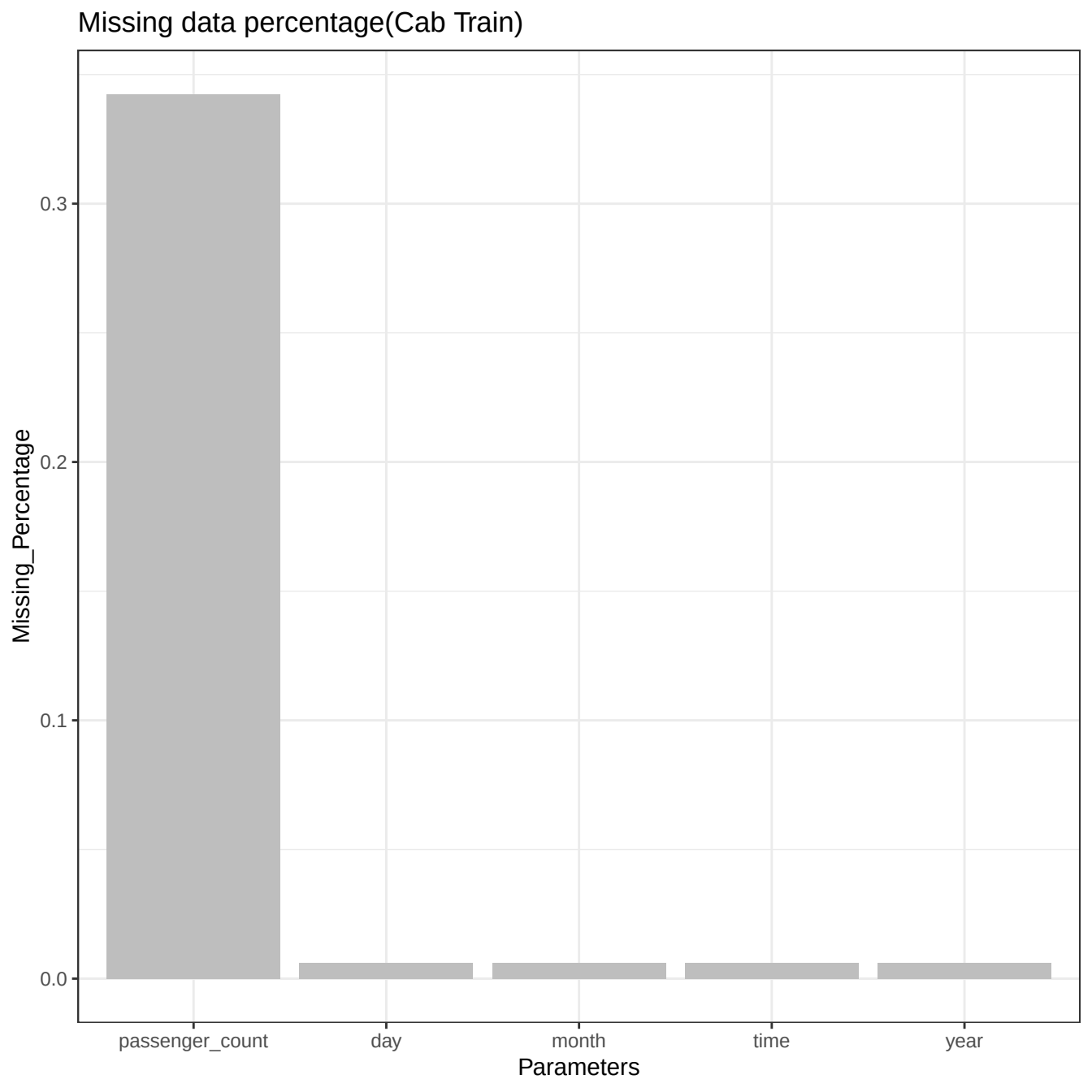
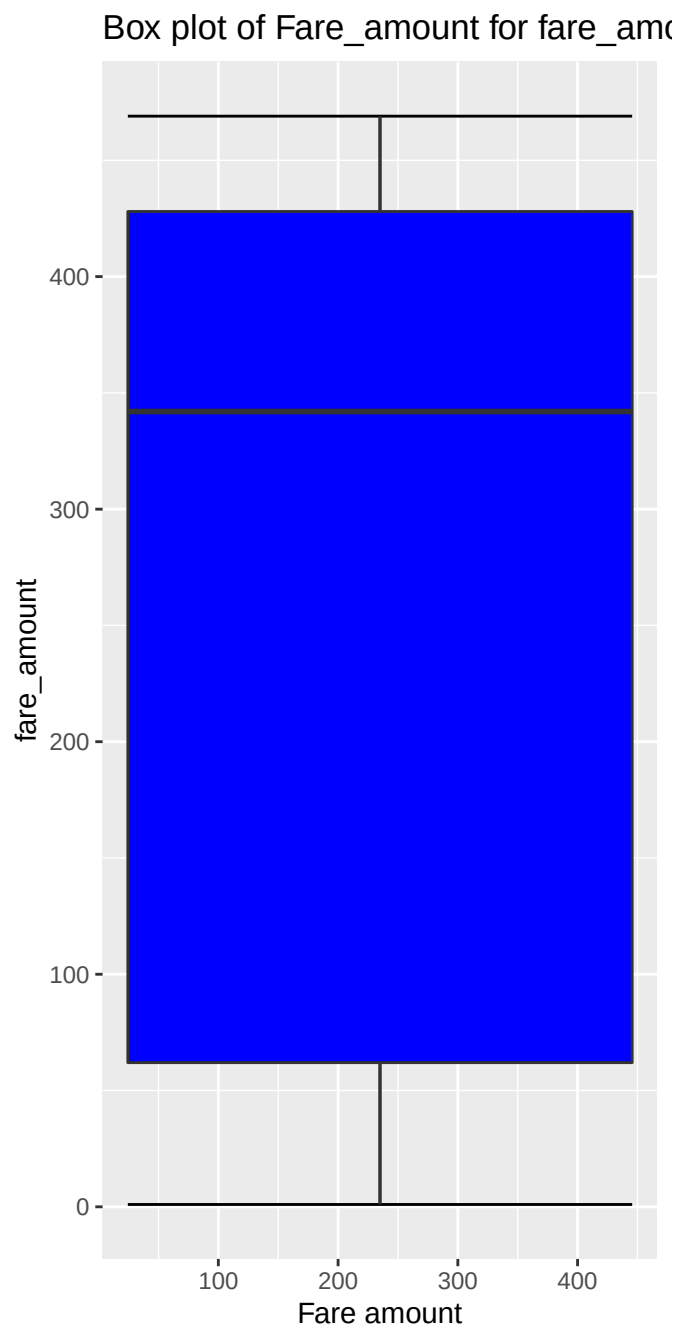
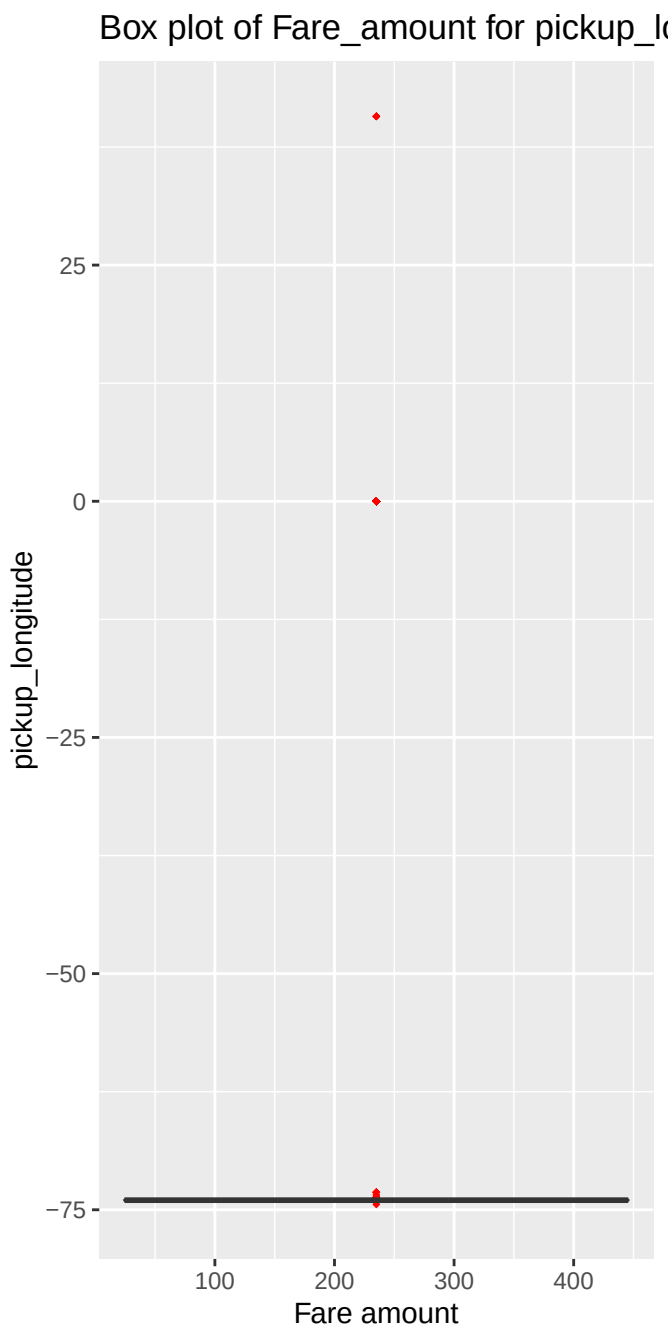
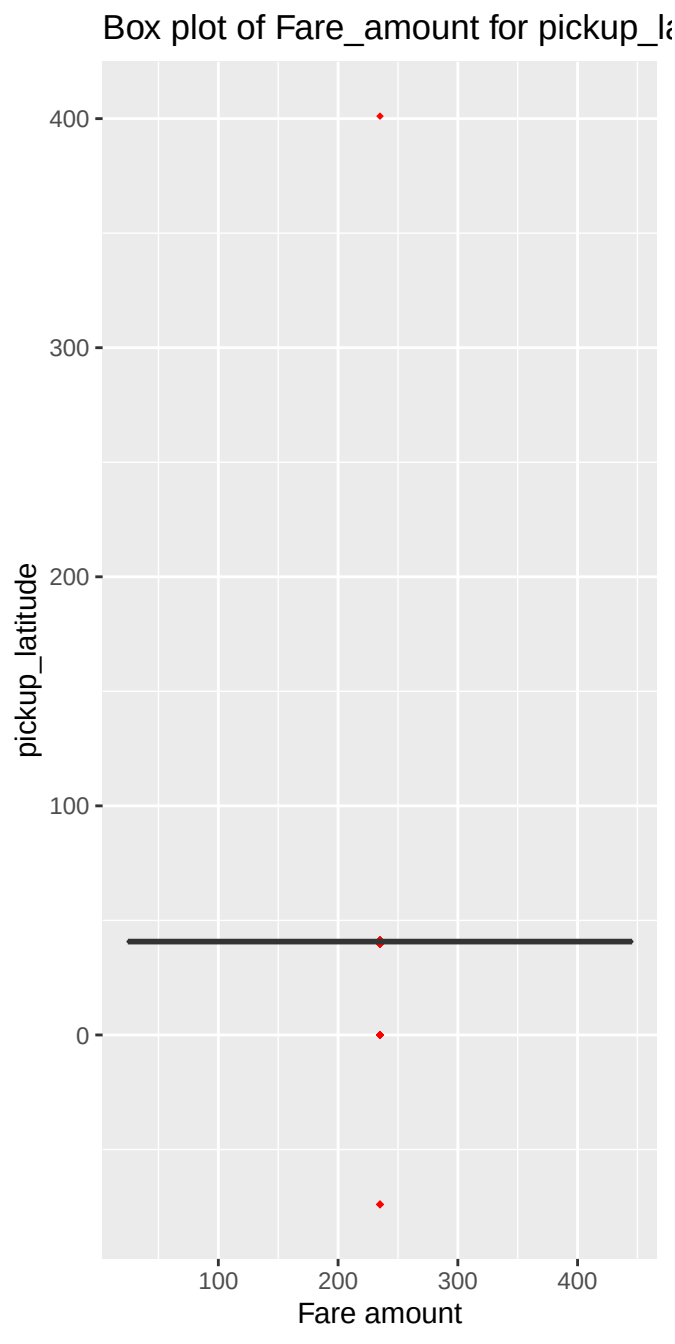
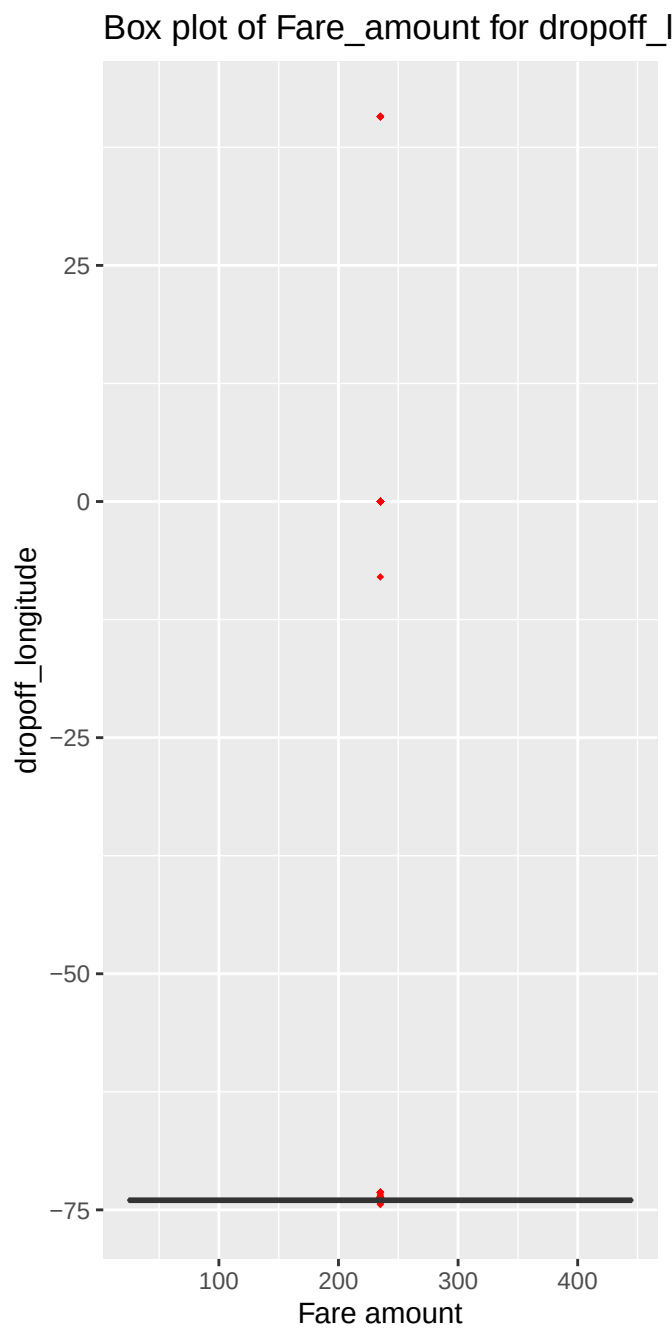
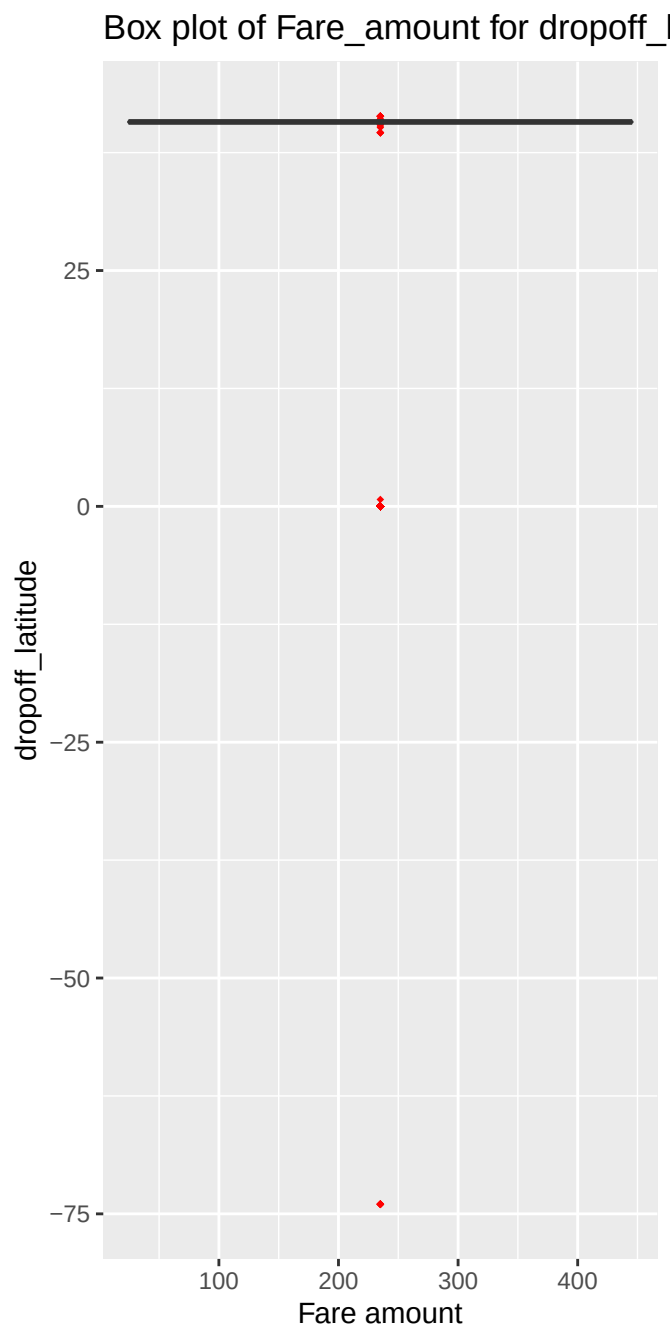

Missing data percentage(Cab Train)
Missing_Percentage
0.3
0.2
0.1
0.0
passenger_count
day
month
time
year
Parameters

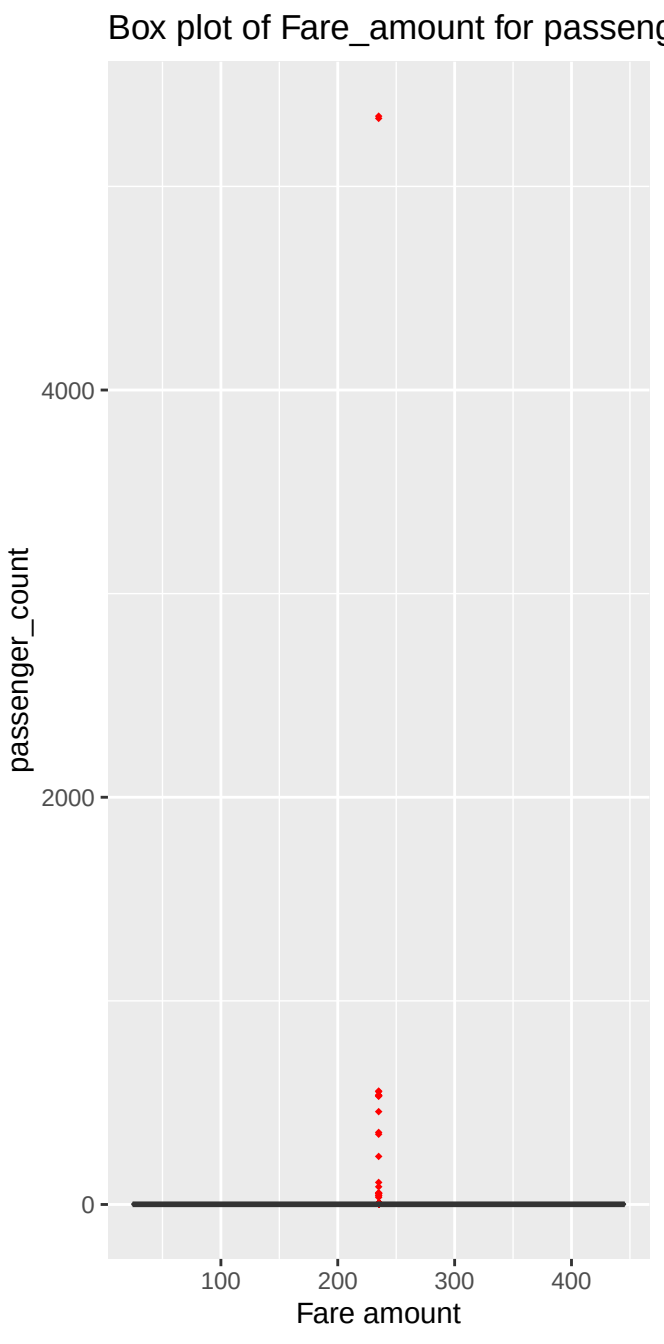

Box plot of Fare_amount for fare_amo
fare_amount
400
300
200
100
0
100
200
300
400
Fare amount

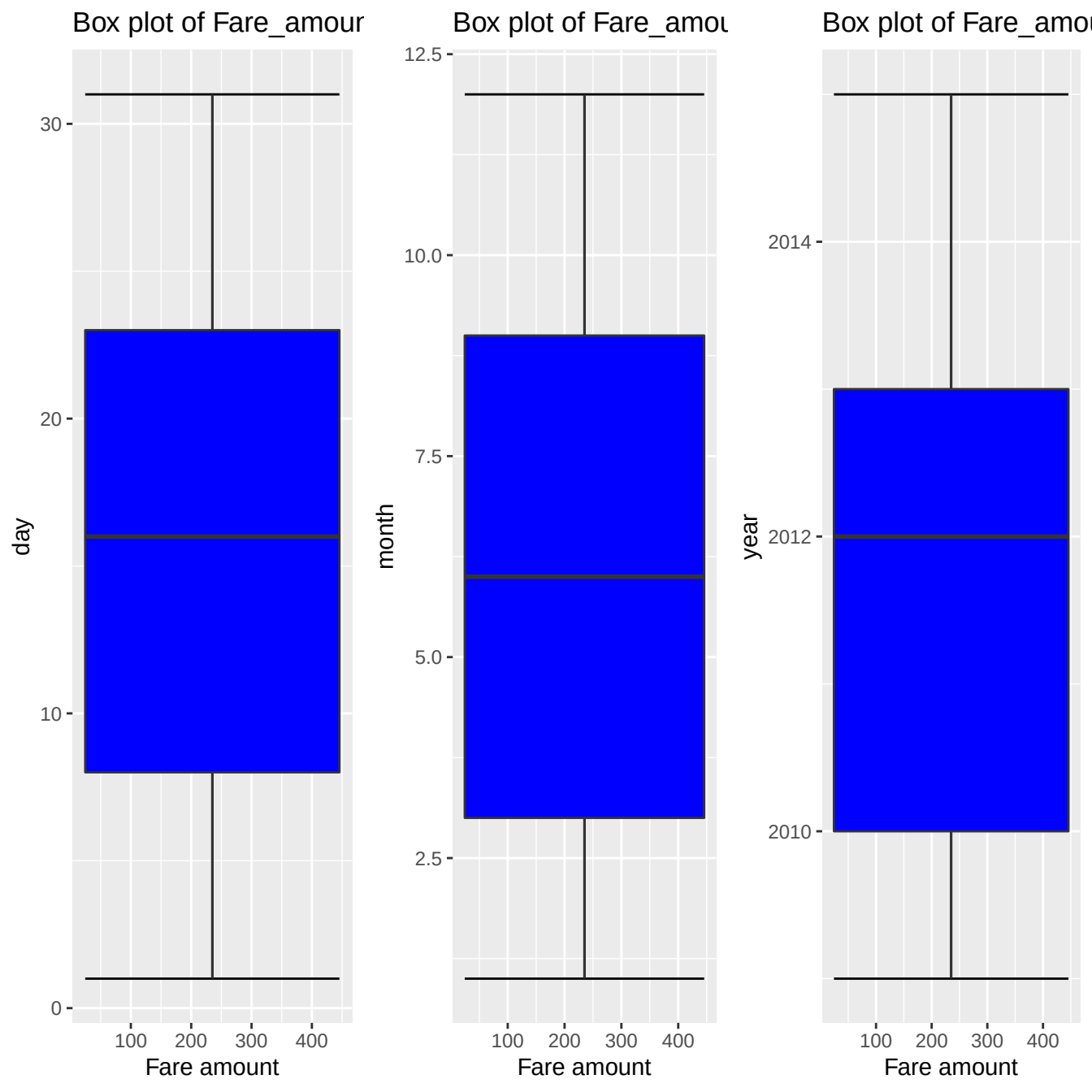

Box plot of Fare_amount for pickup_lo
pickup_longitude
25
0
−25
−50
−75
100
200
300
400
Fare amount

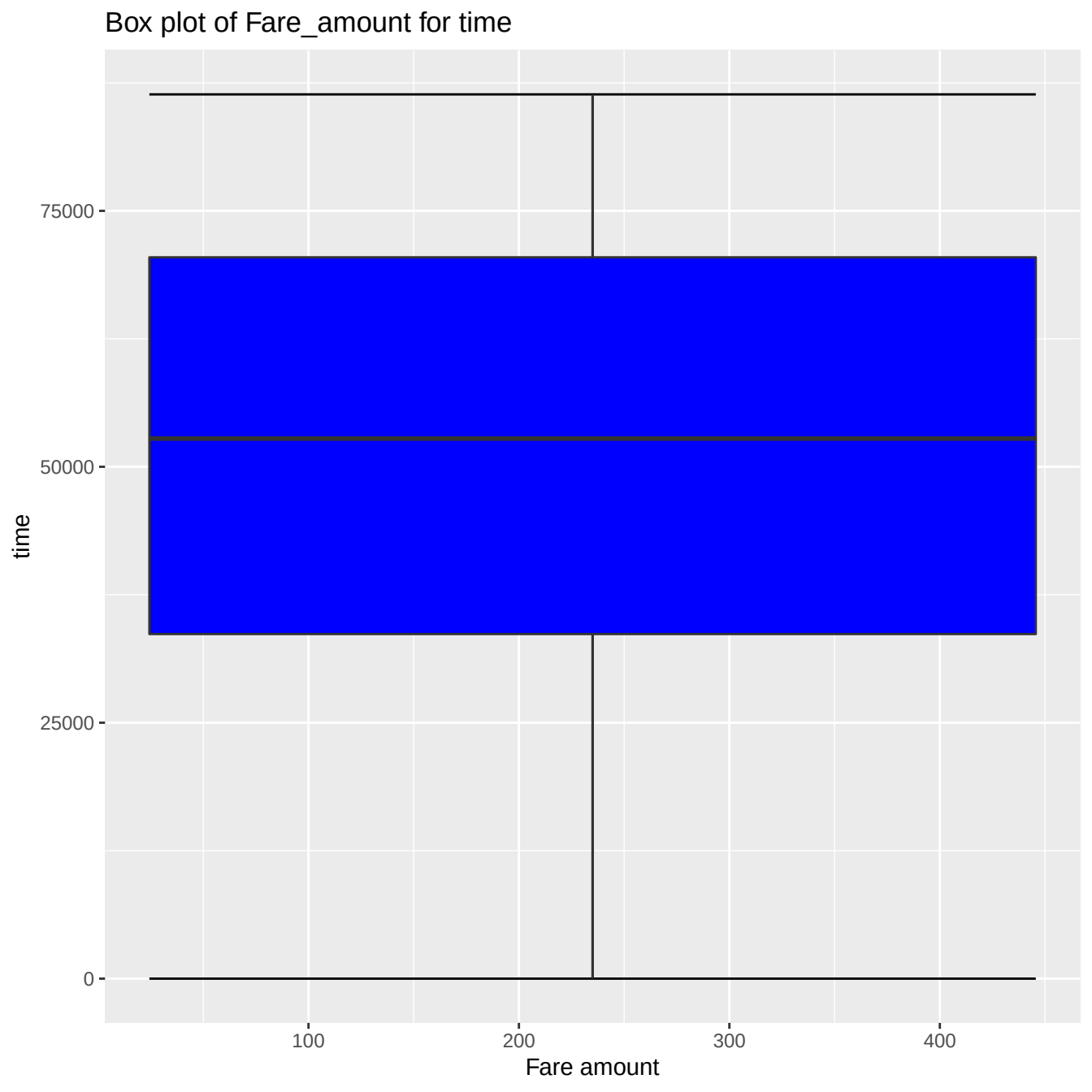

Box plot of Fare_amount for pickup_la
400
300
200
100
0
pickup_latitude
100
200
300
400
Fare amount


Box plot of Fare_amount for dropoff_l
25
0
−25
−50
−75
dropoff_longitude
100
200
300
400
Fare amount

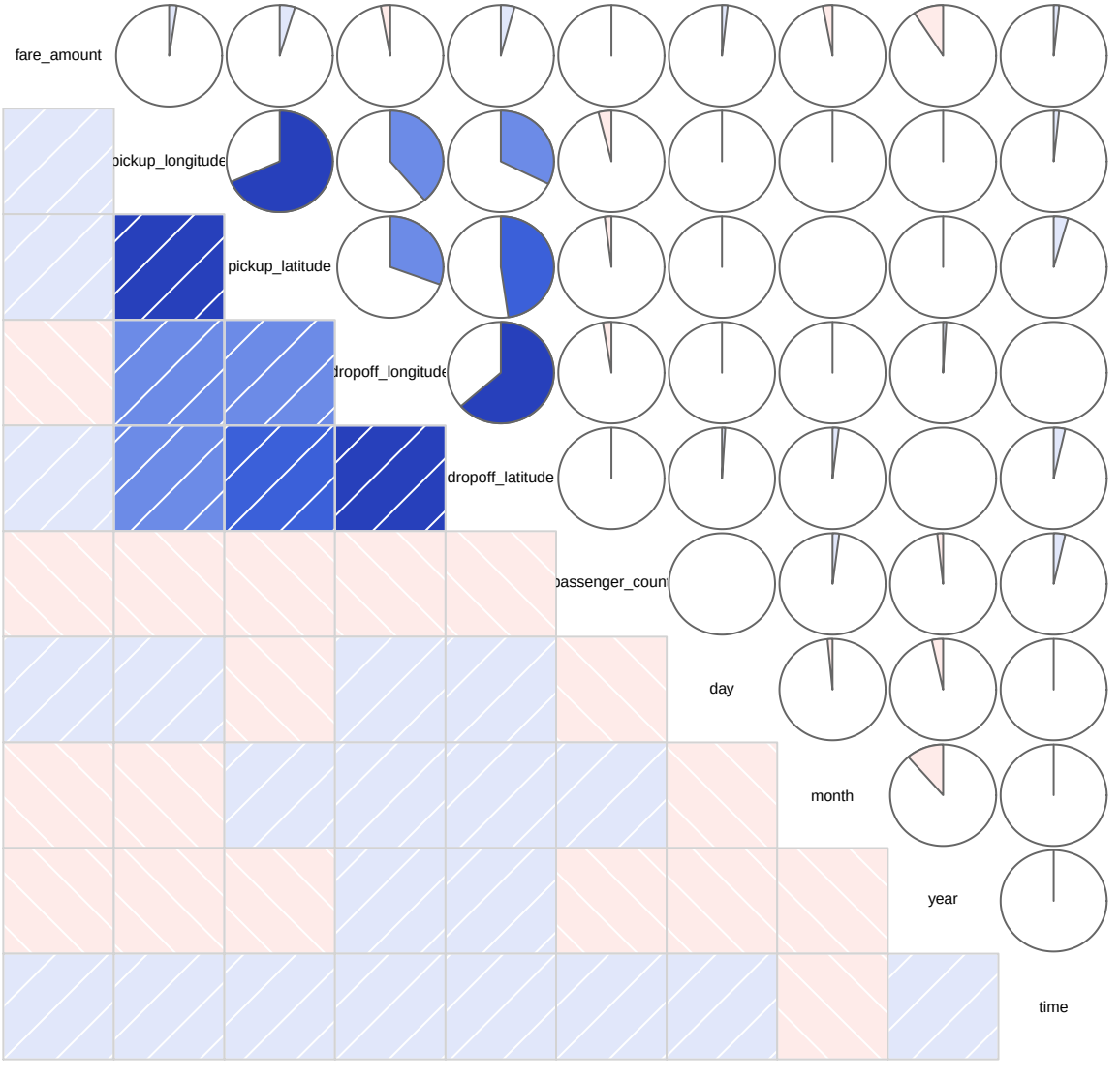

Box plot of Fare_amount for dropoff_l
25
0
−25
−50
−75
dropoff_latitude
100
200
300
400
Fare amount

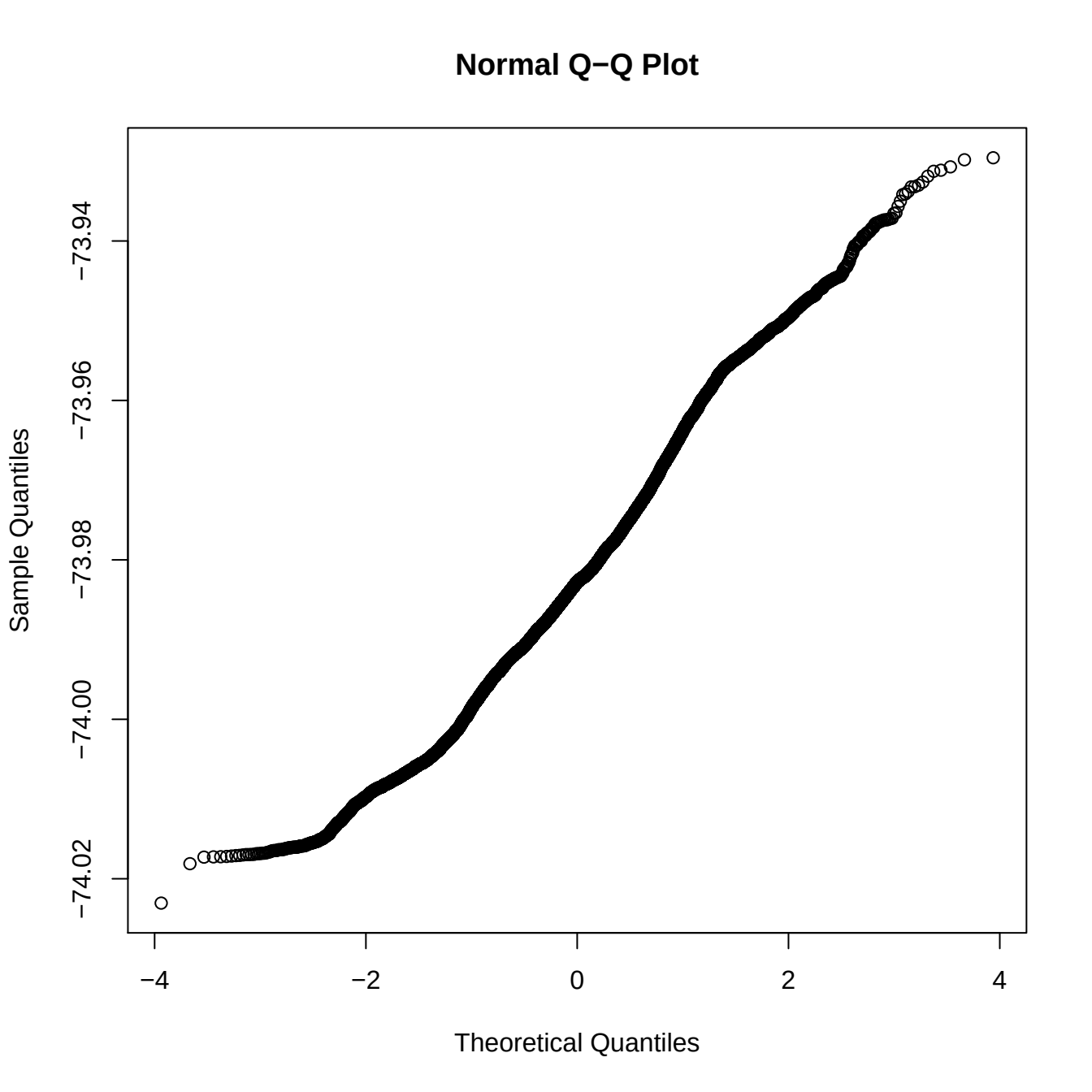

Box plot of Fare_amount for passeng
4000
2000
0
passenger_count
100
200
300
400
Fare amount

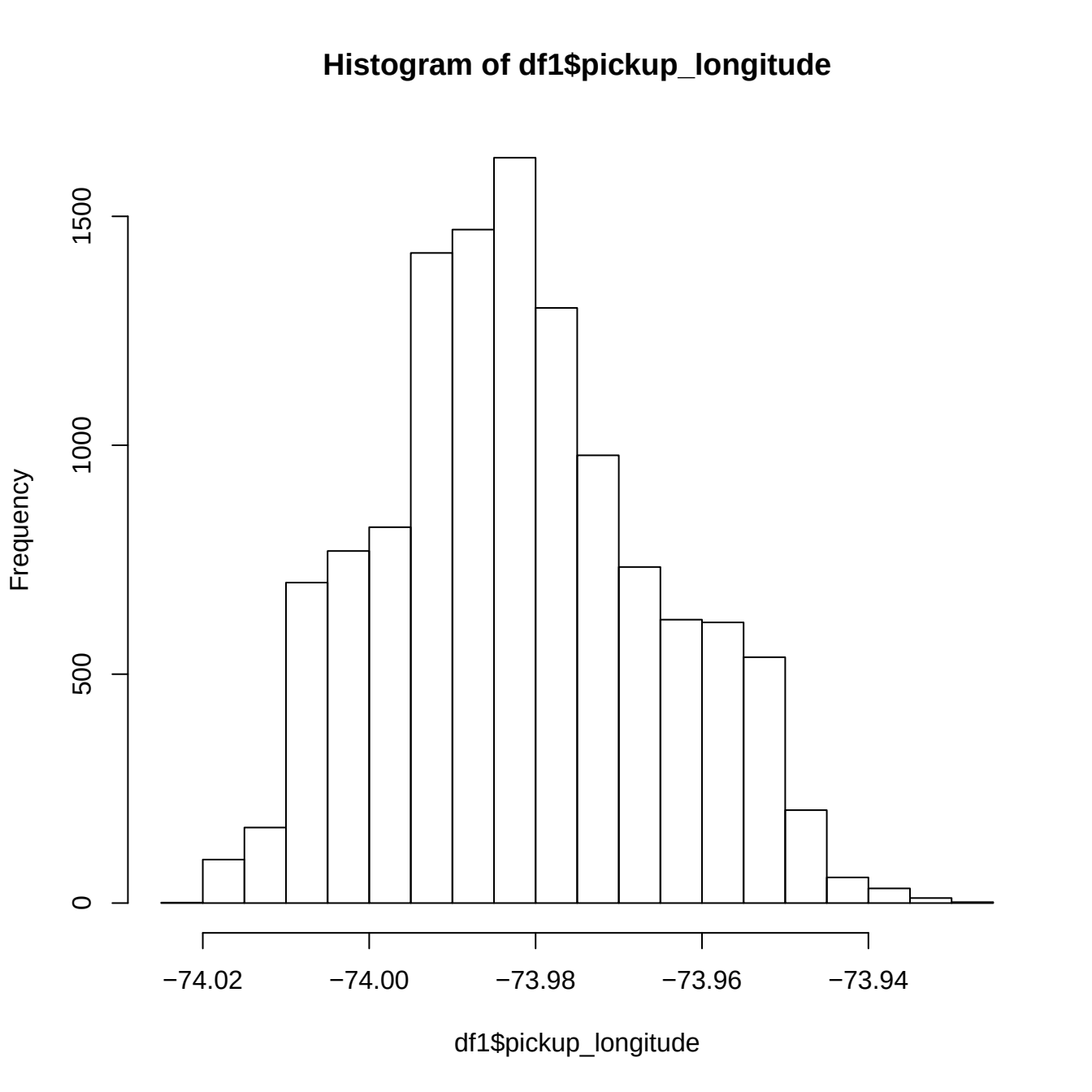

Box plot of Fare_amour
day
30
20
10
0
100
200
300
400
Fare amount
Box plot of Fare_amou
month
12.5
10.0
7.5
5.0
2.5
100
200
300
400
Fare amount
Box plot of Fare_amou
year
2014
2012
2010
100
200
300
400
Fare amount

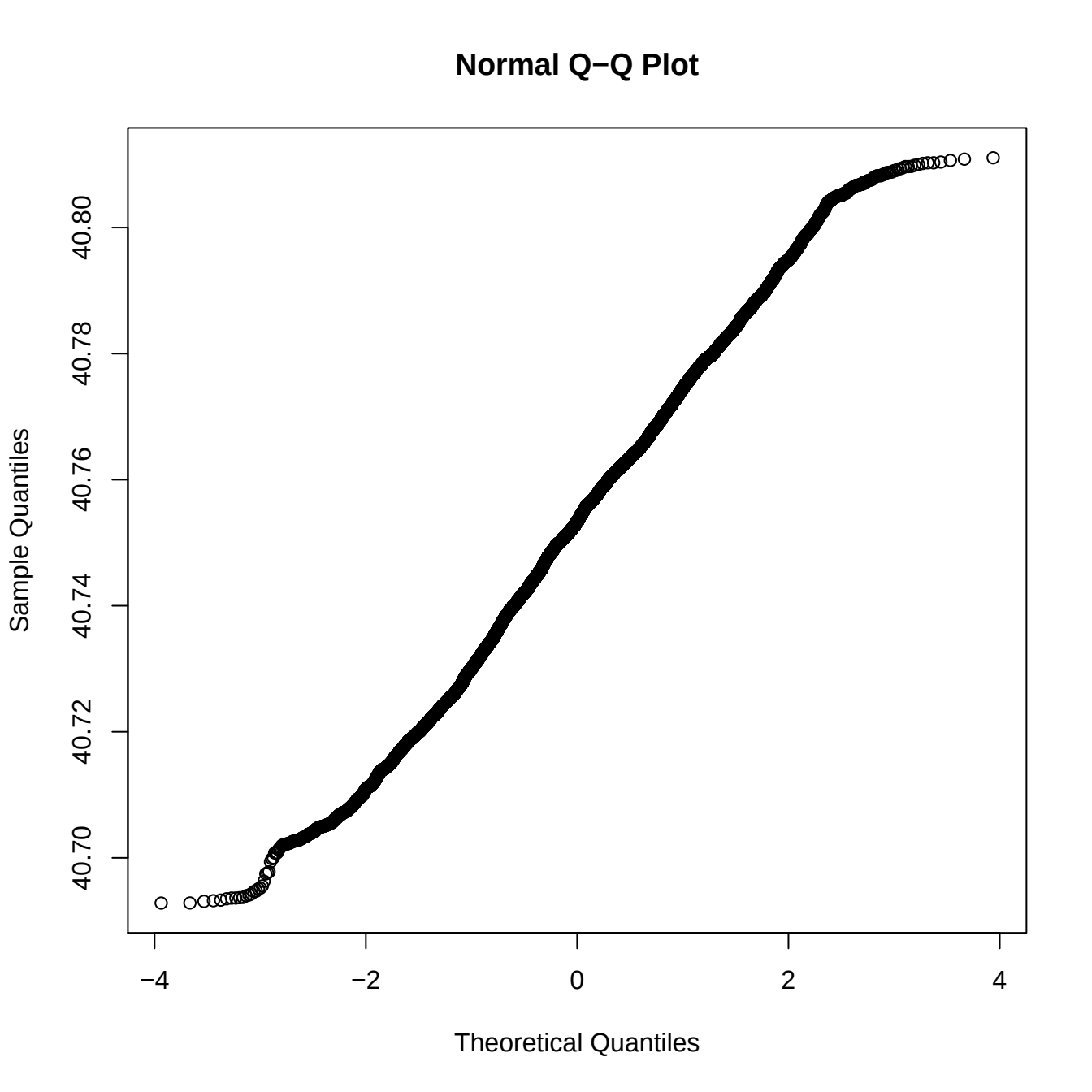

Box plot of Fare_amount for time
75000
50000
25000
0
time
100
200
300
400
Fare amount

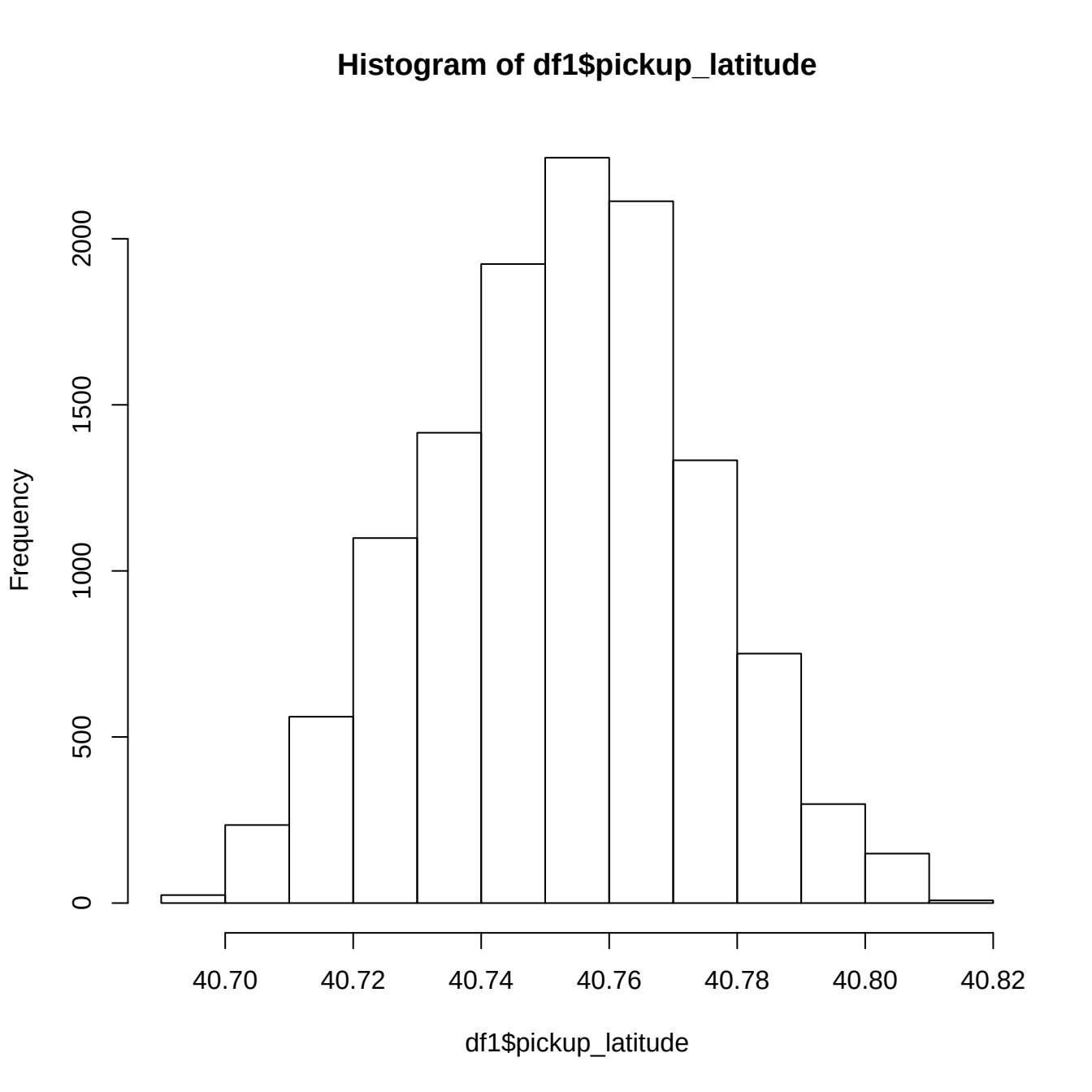
Correlation Plot

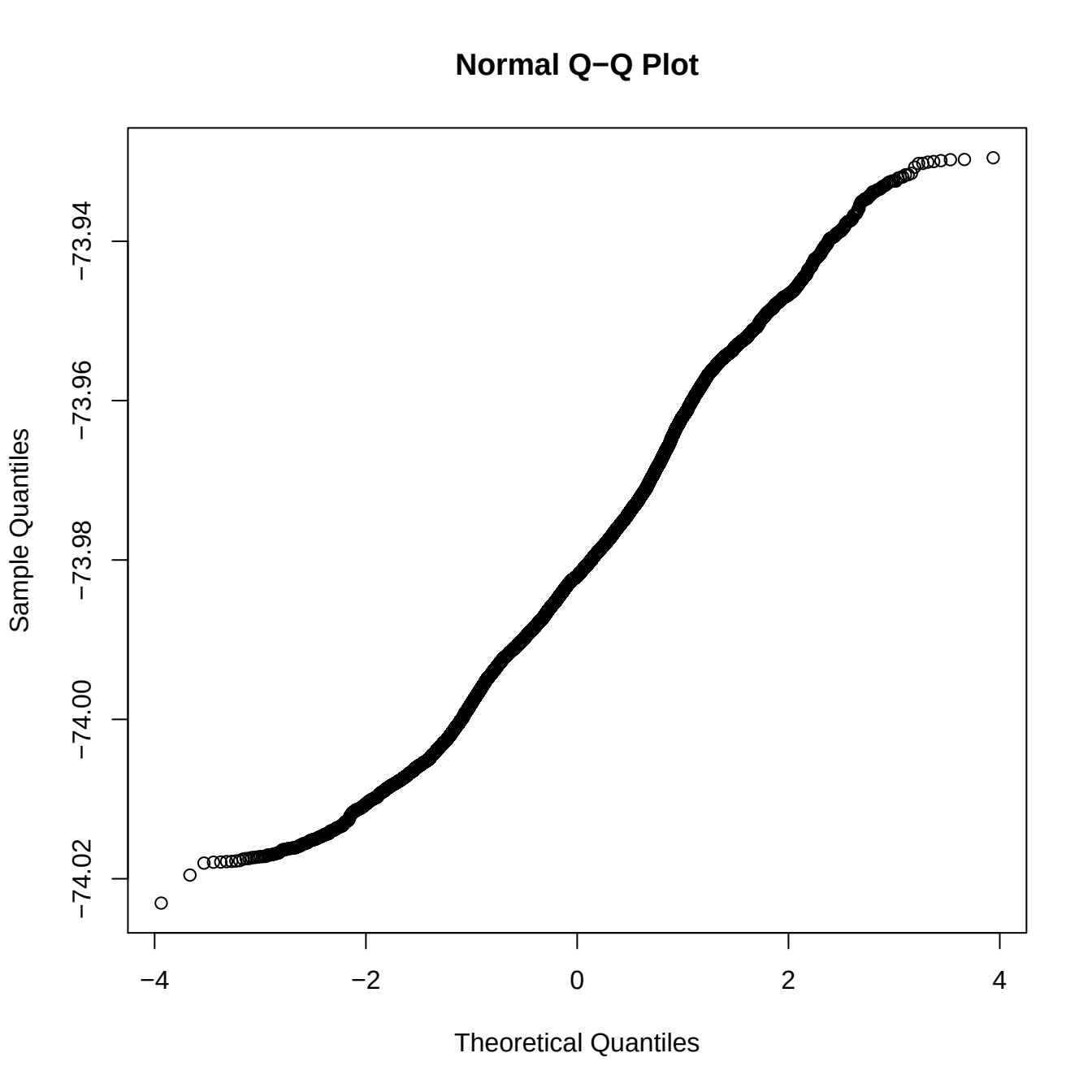


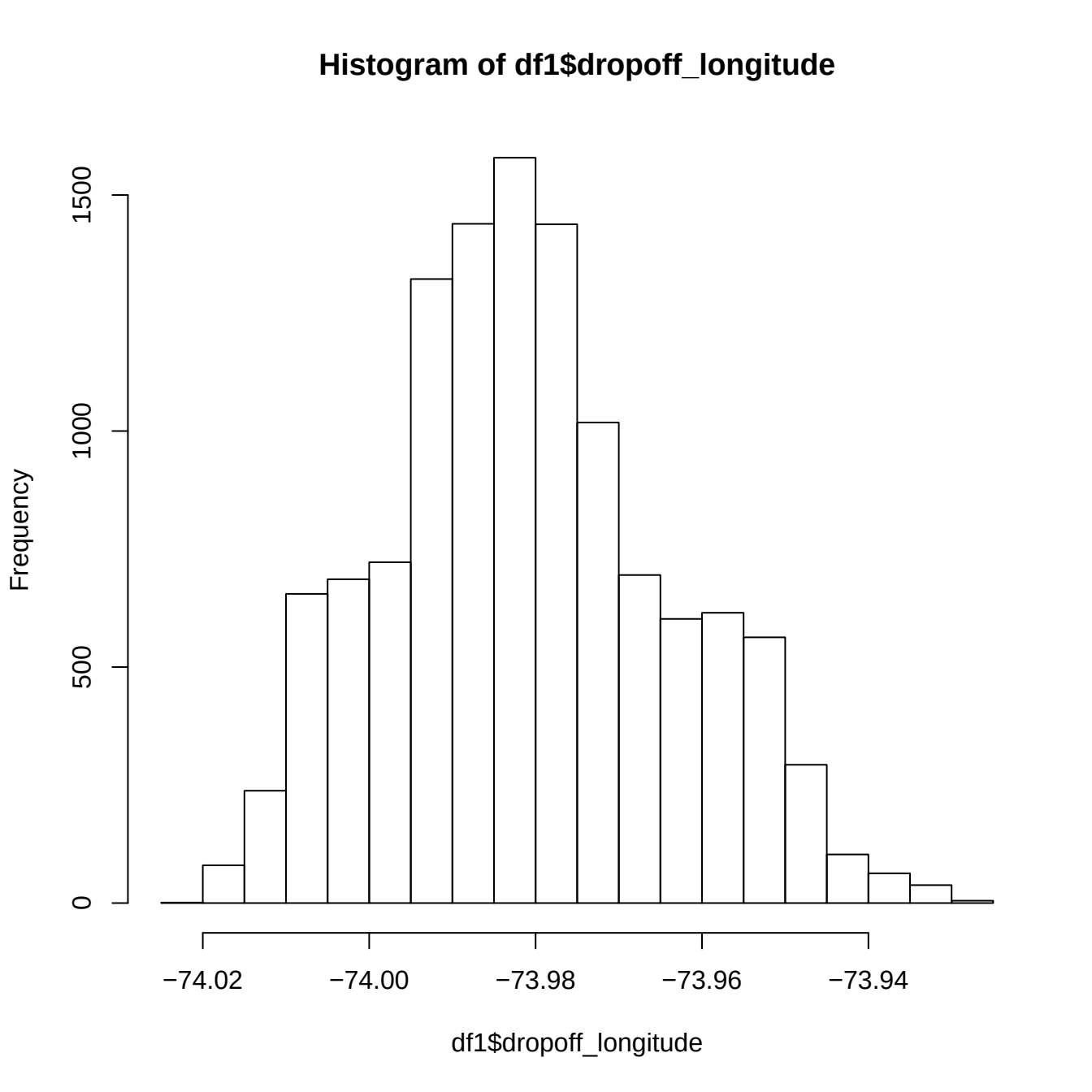

Normal Q−Q Plot
Sample Quantiles
−74.02
−74.00
−73.98
−73.96
−73.94
Theoretical Quantiles
−4
−2
0
2
4

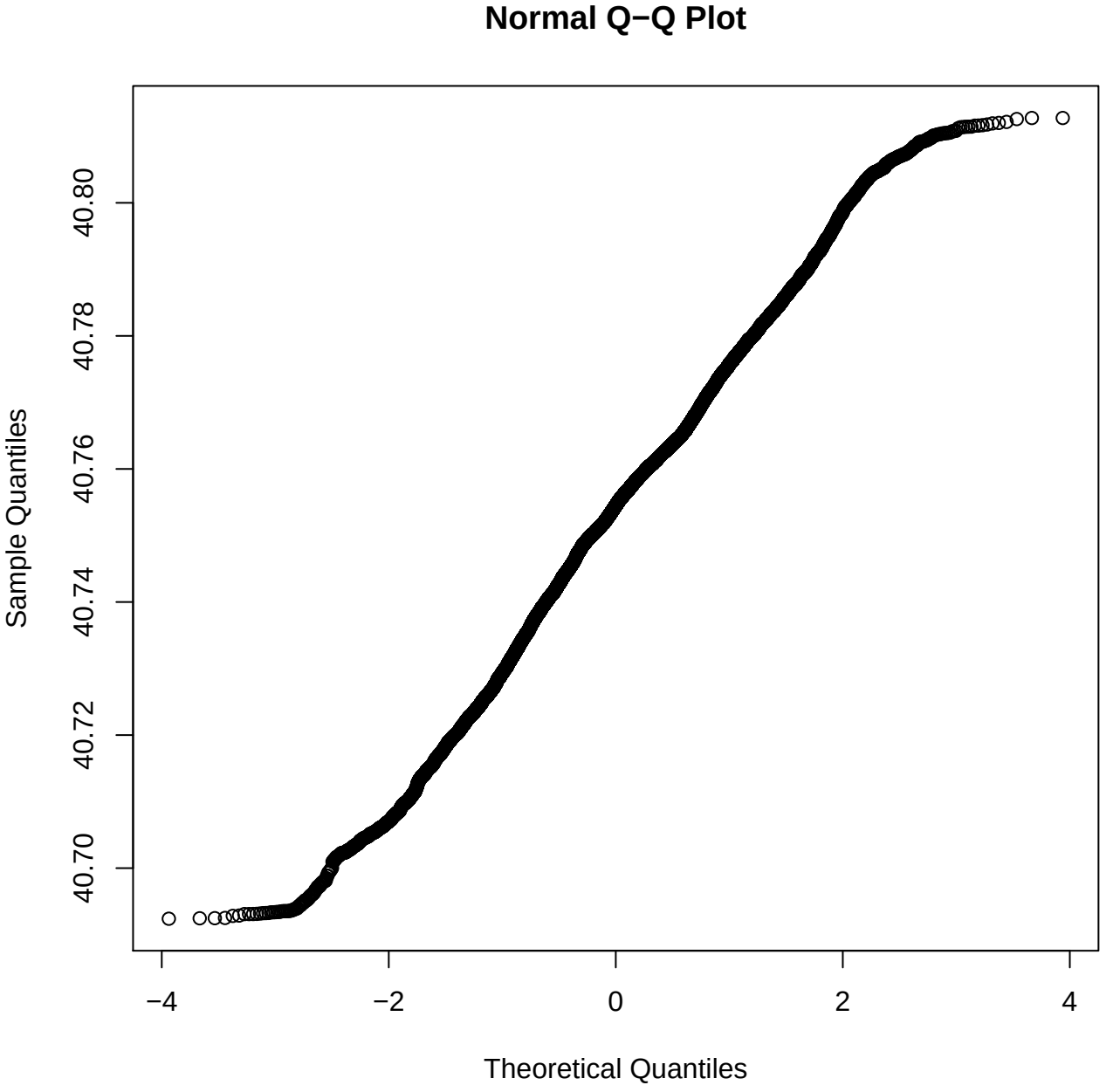

Histogram of df1$pickup_longitude
Frequency
0
500
1000
1500
−74.02
−74.00
−73.98
−73.96
−73.94
df1$pickup_longitude

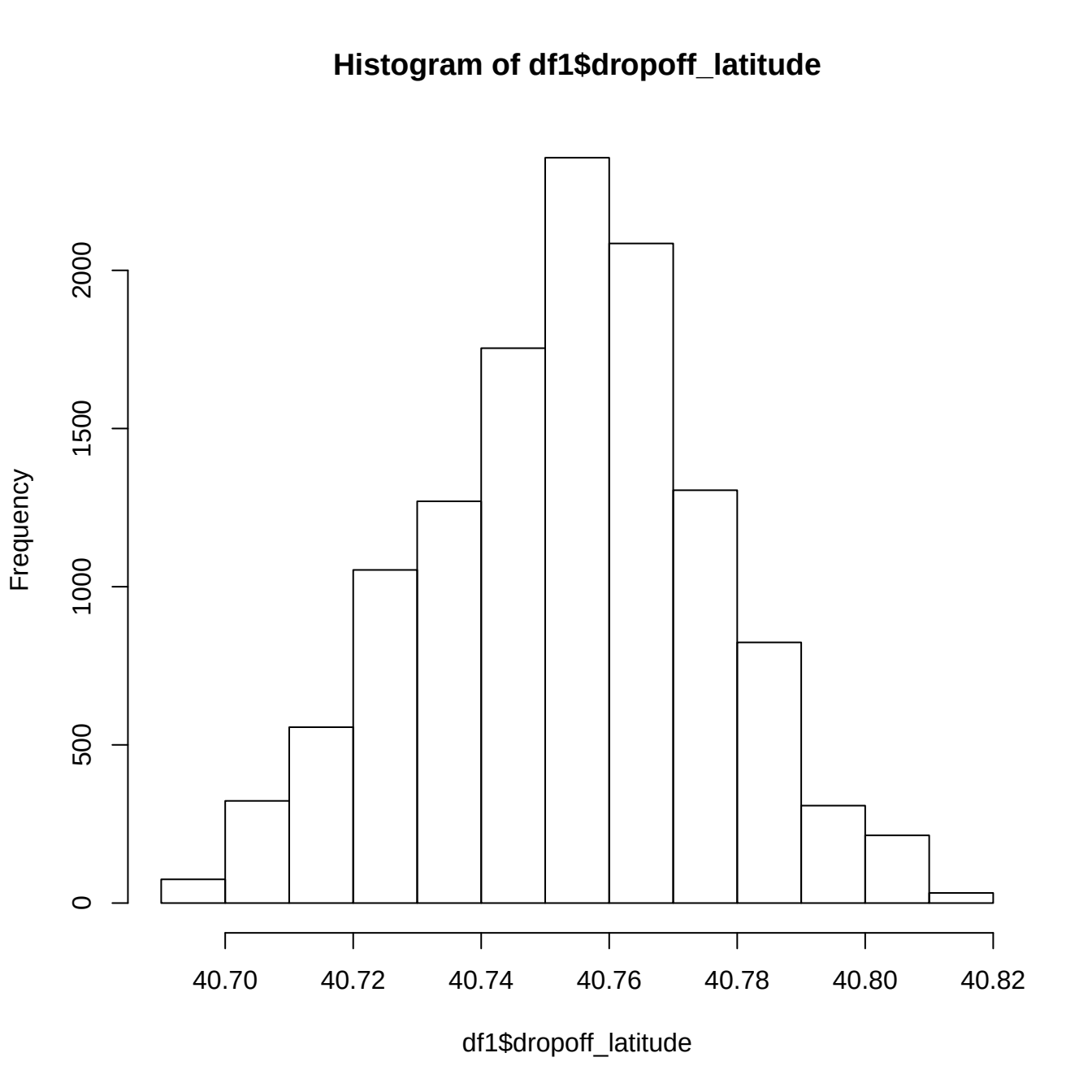

Normal Q−Q Plot
Sample Quantiles
40.70
40.72
40.74
40.76
40.78
40.80
−4
−2
0
2
4
Theoretical Quantiles

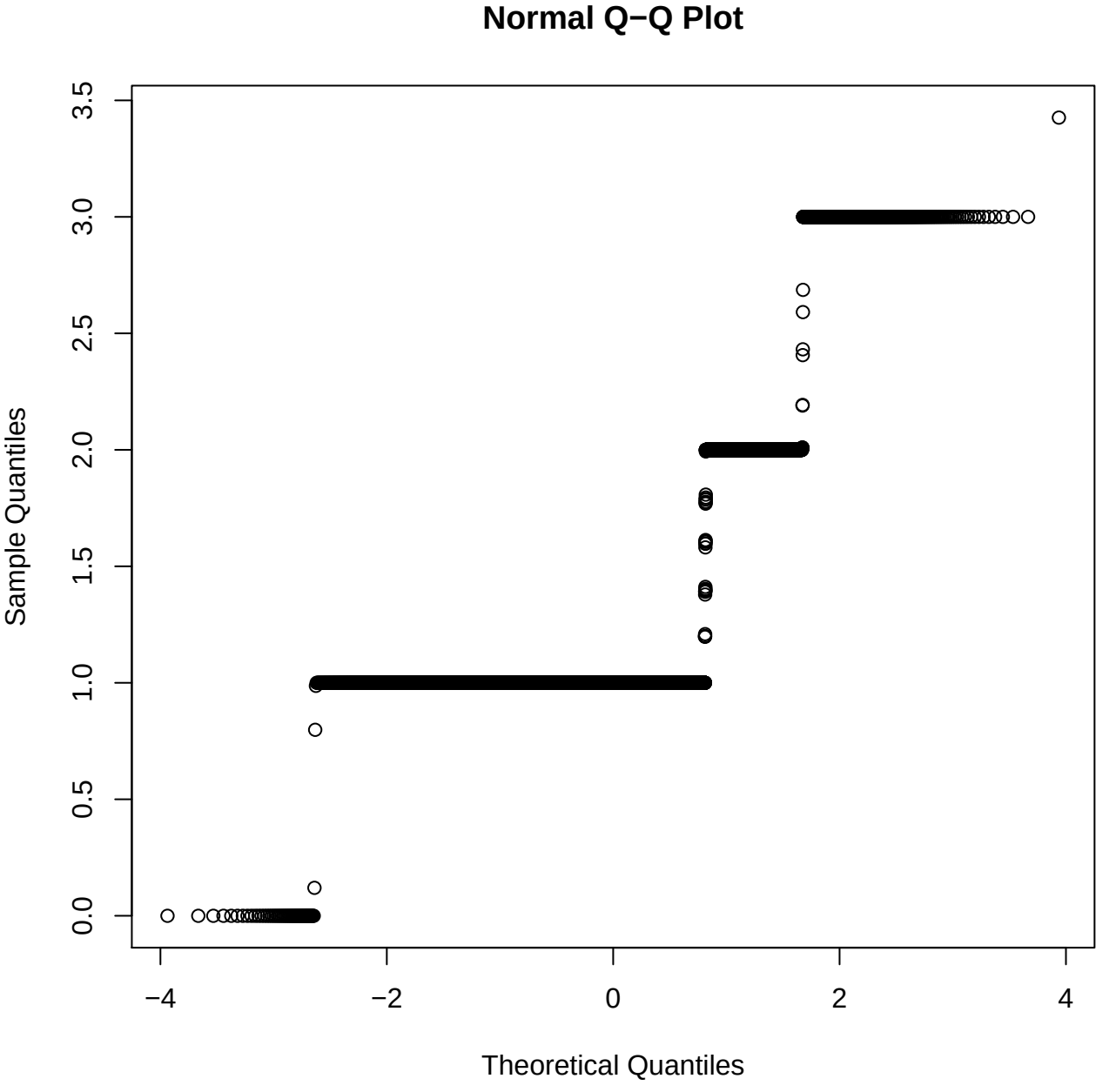

Histogram of df1$pickup_latitude
Frequency
0
500
1000
1500
2000
40.70
40.72
40.74
40.76
40.78
40.80
40.82
df1$pickup_latitude

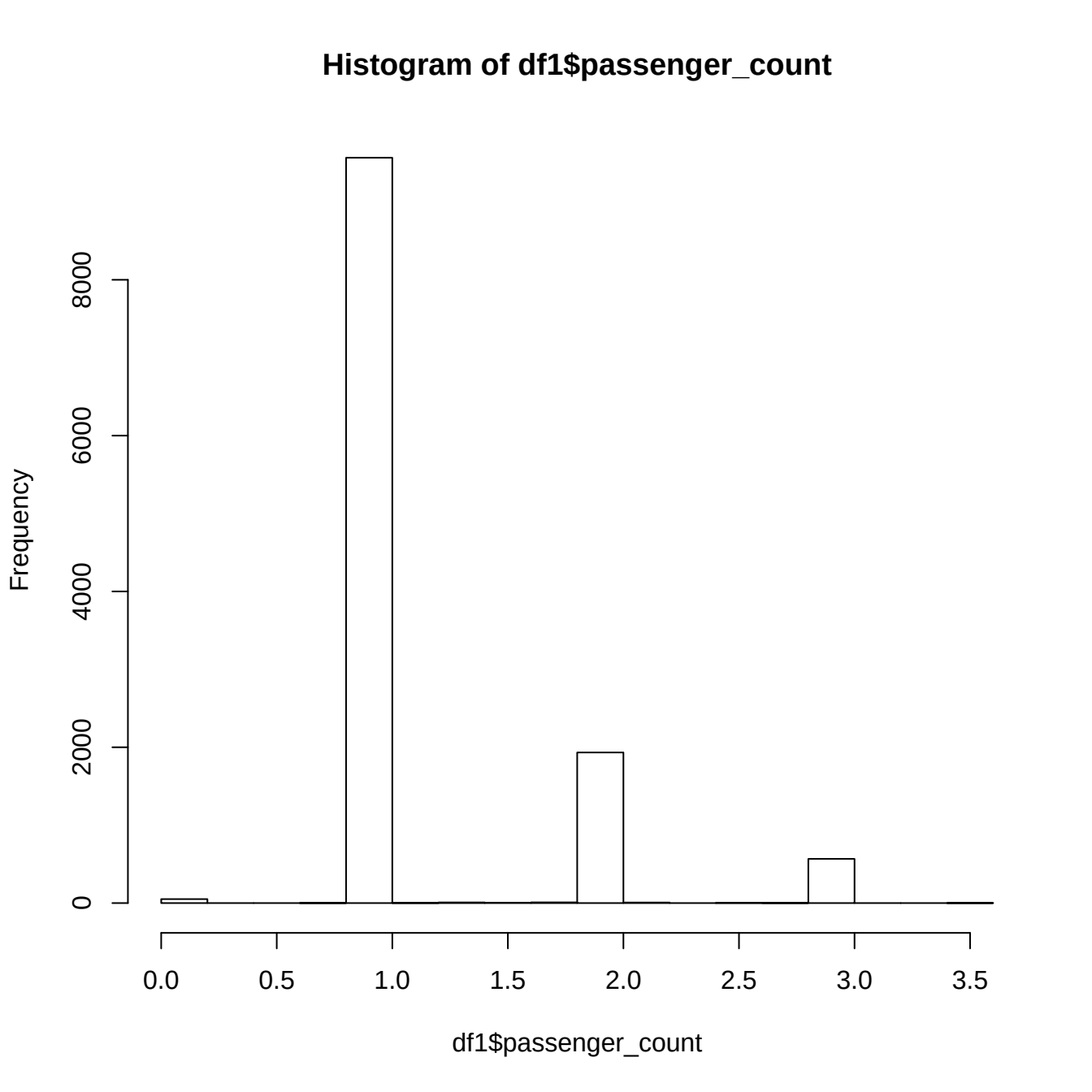

Normal Q−Q Plot
Sample Quantiles
−74.02
−74.00
−73.98
−73.96
−73.94
Theoretical Quantiles
−4
−2
0
2
4

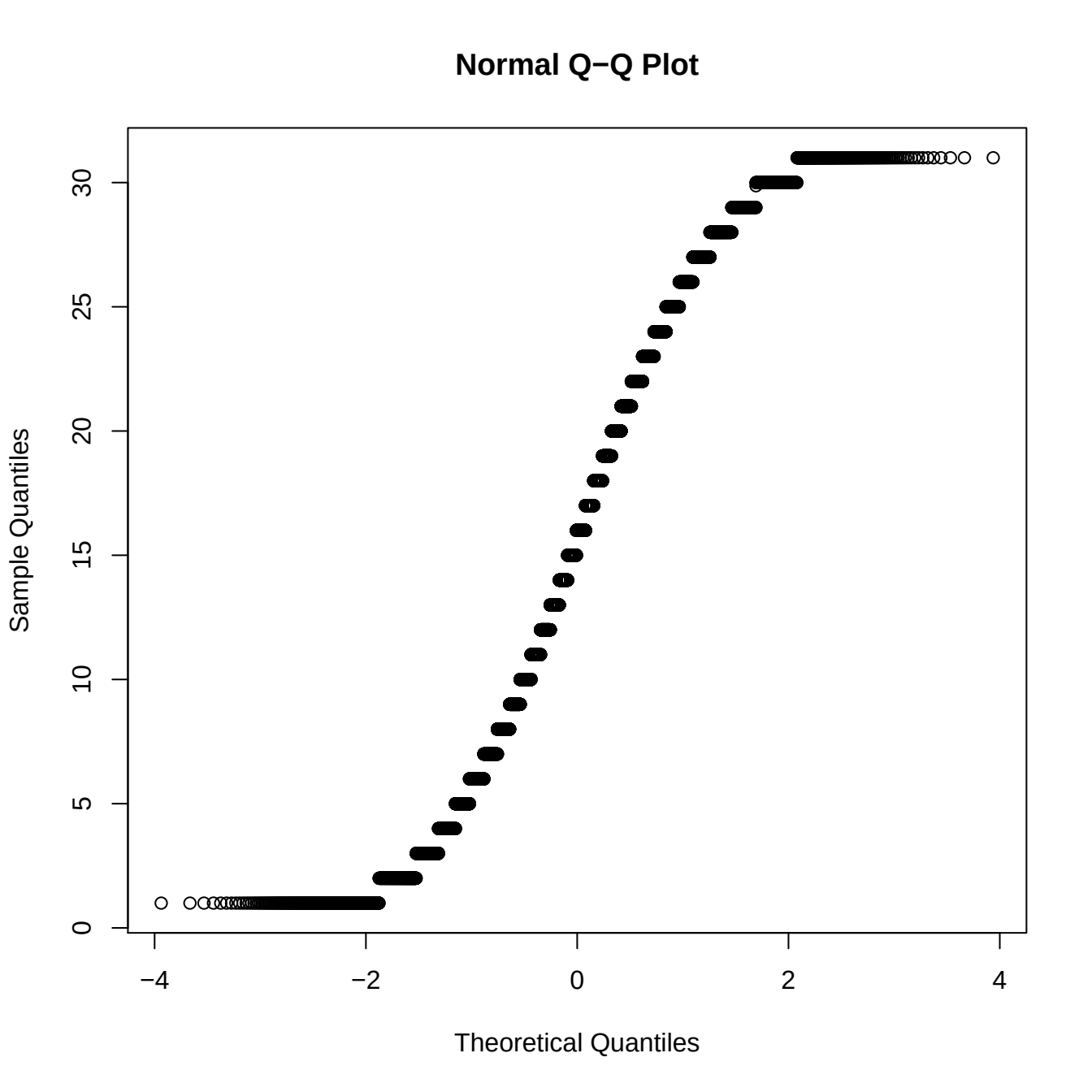

Histogram of df1$dropoff_longitude
Frequency
0
500
1000
1500
−74.02
−74.00
−73.98
−73.96
−73.94
df1$dropoff_longitude

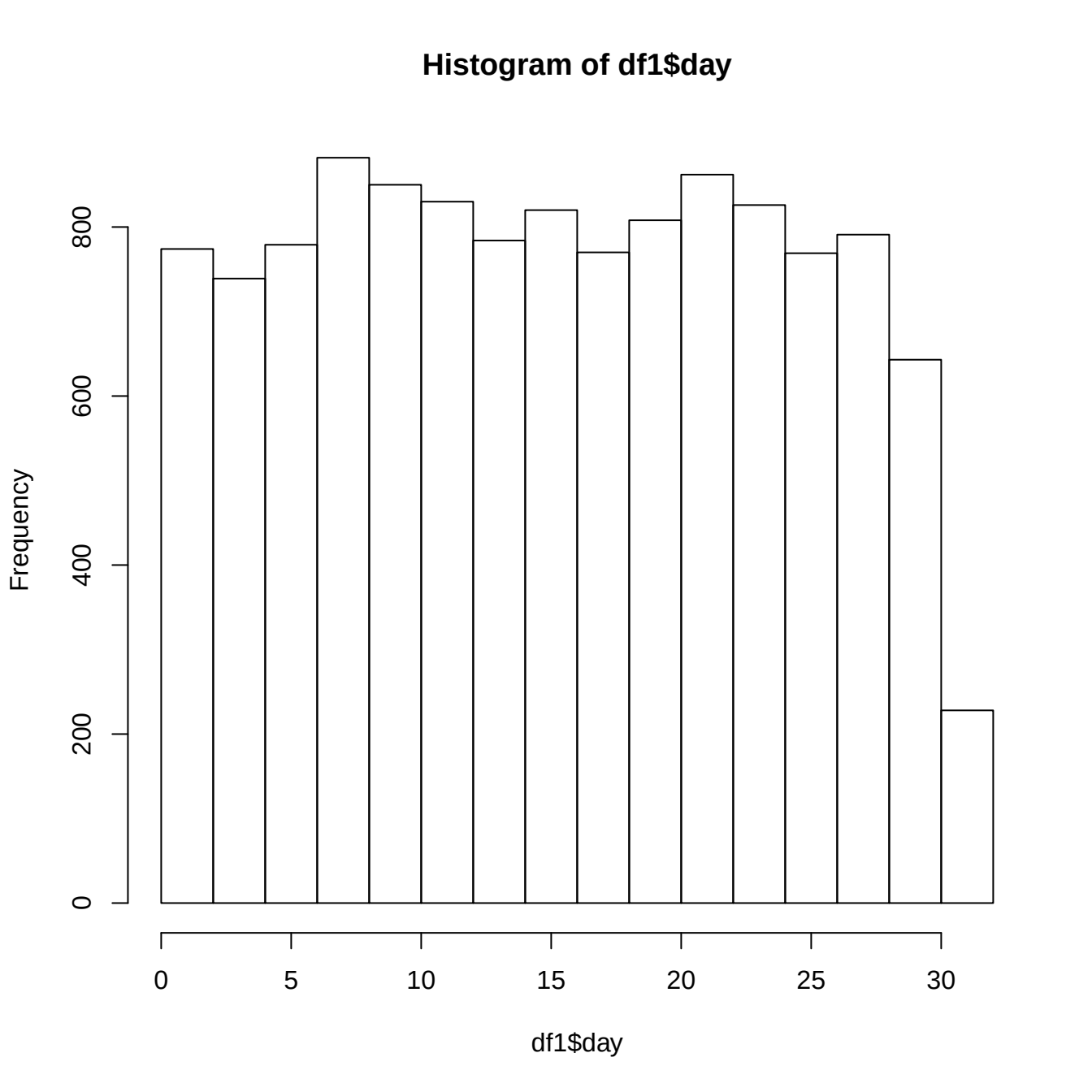

Normal Q−Q Plot
Sample Quantiles
40.70
40.72
40.74
40.76
40.78
40.80
−4
−2
0
2
4
Theoretical Quantiles

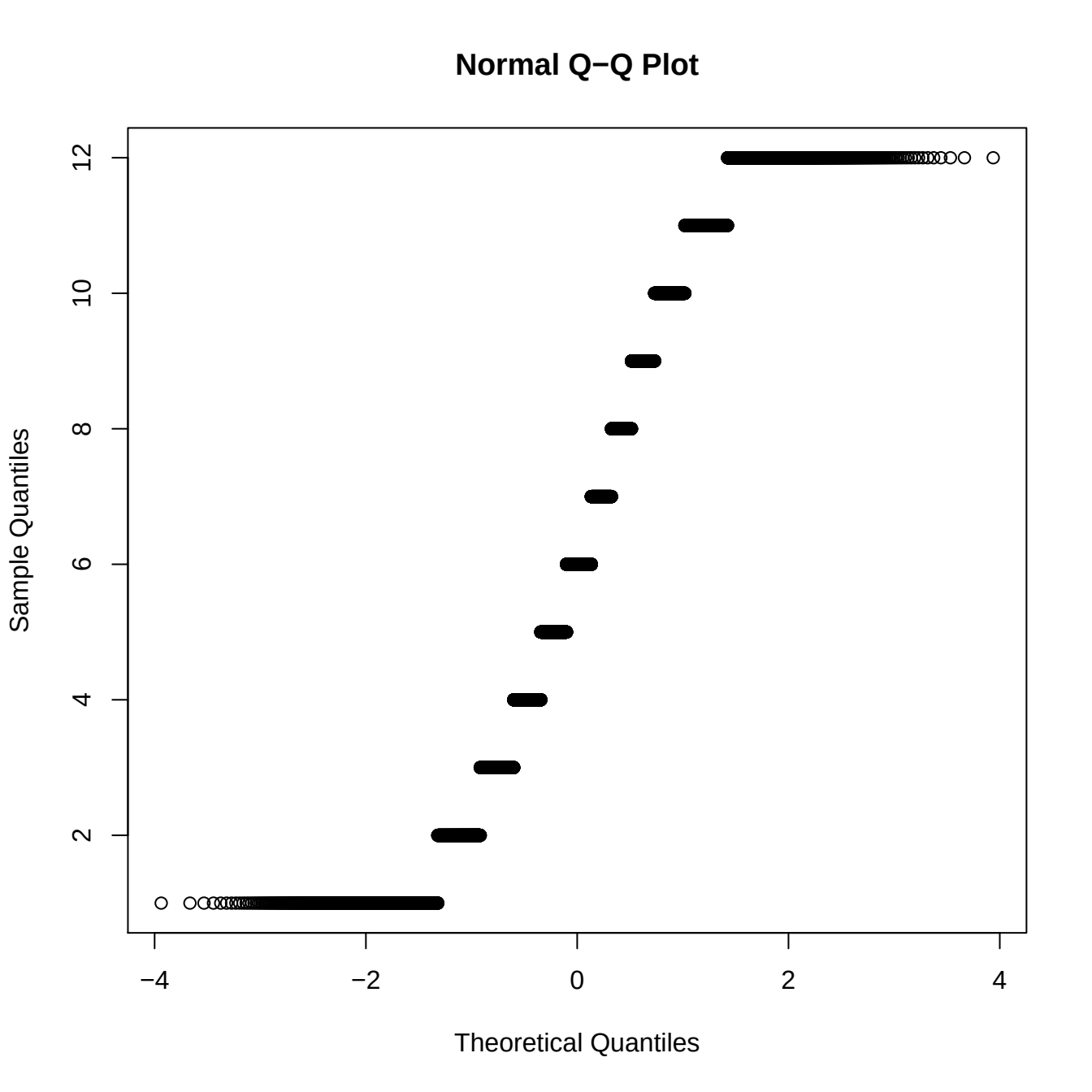

Histogram of df1$dropoff_latitude
Frequency
0
500
1000
1500
2000
40.70
40.72
40.74
40.76
40.78
40.80
40.82
df1$dropoff_latitude

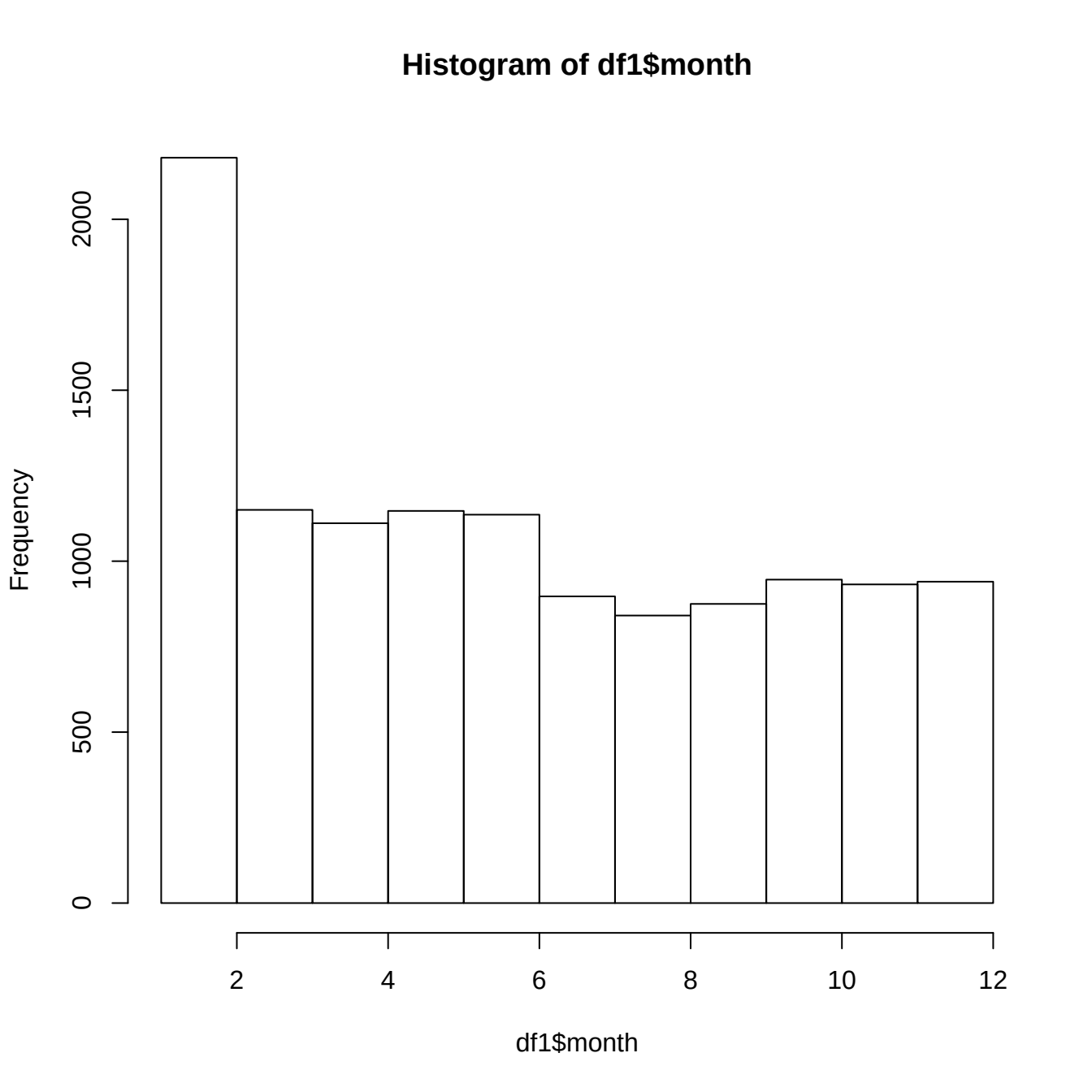

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
−4
−2
0
2
4

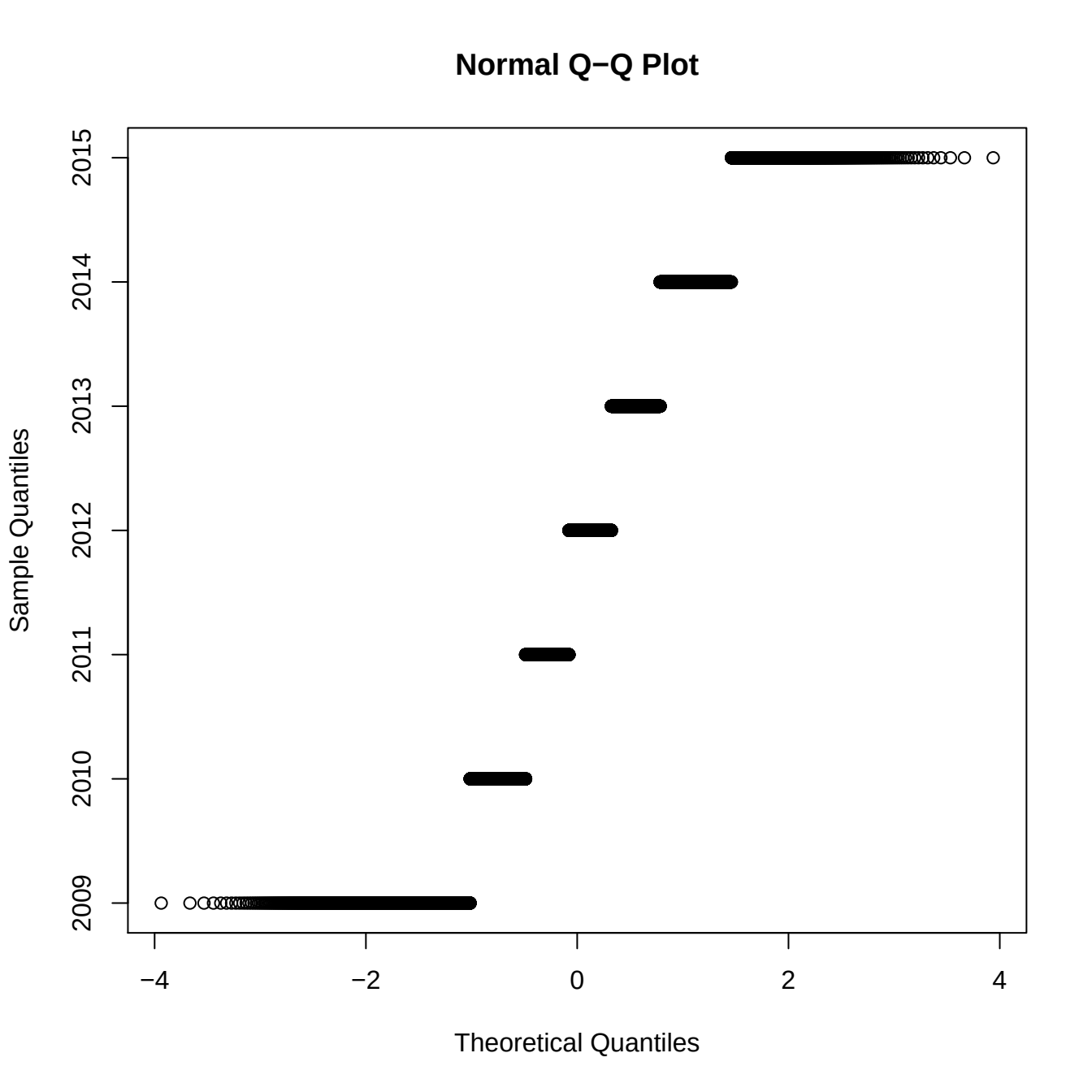

Histogram of df1$passenger_count
Frequency
0
2000
4000
6000
8000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
df1$passenger_count

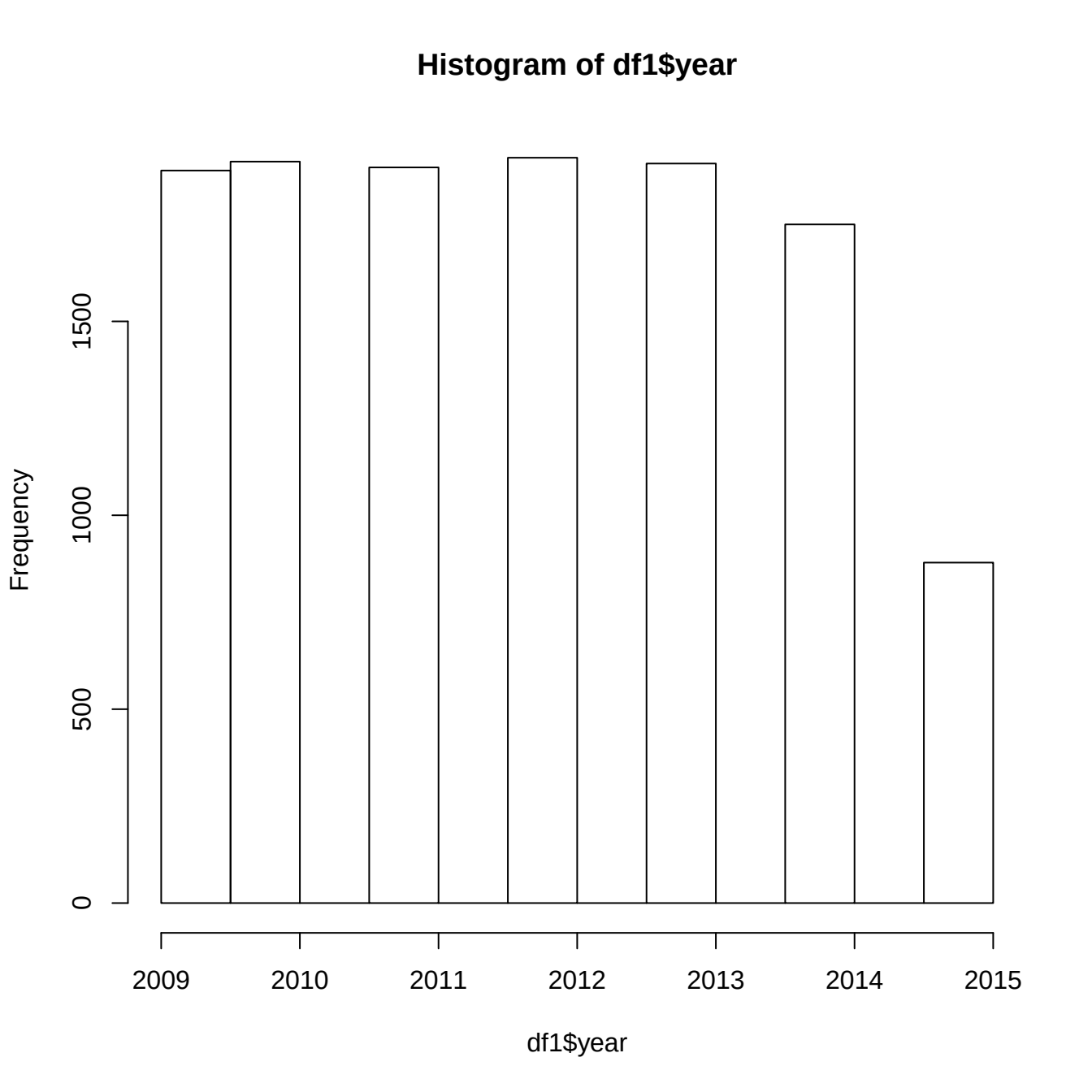

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0
5
10
15
20
25
30
−4
−2
0
2
4

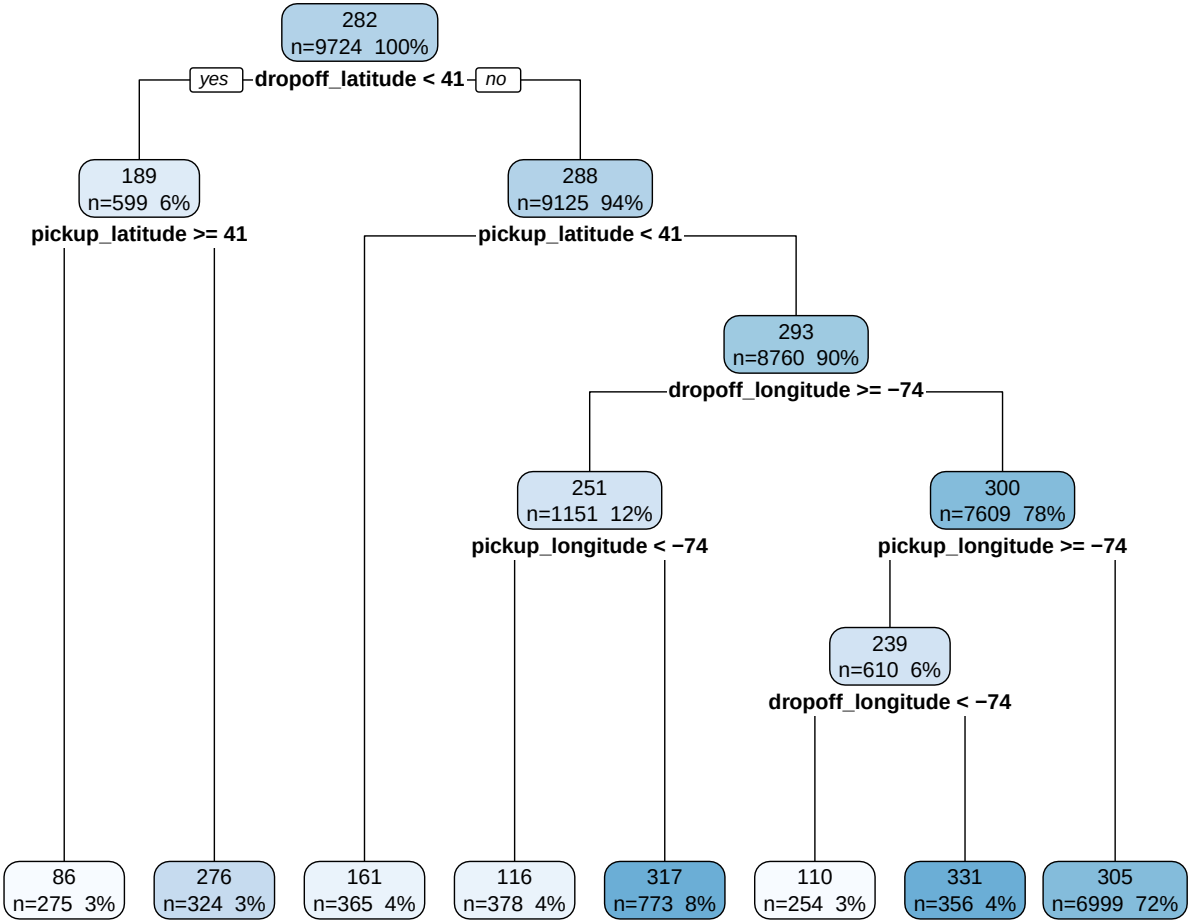

Histogram of df1$day
Frequency
0
200
400
600
800
0
5
10
15
20
25
30
df1$day

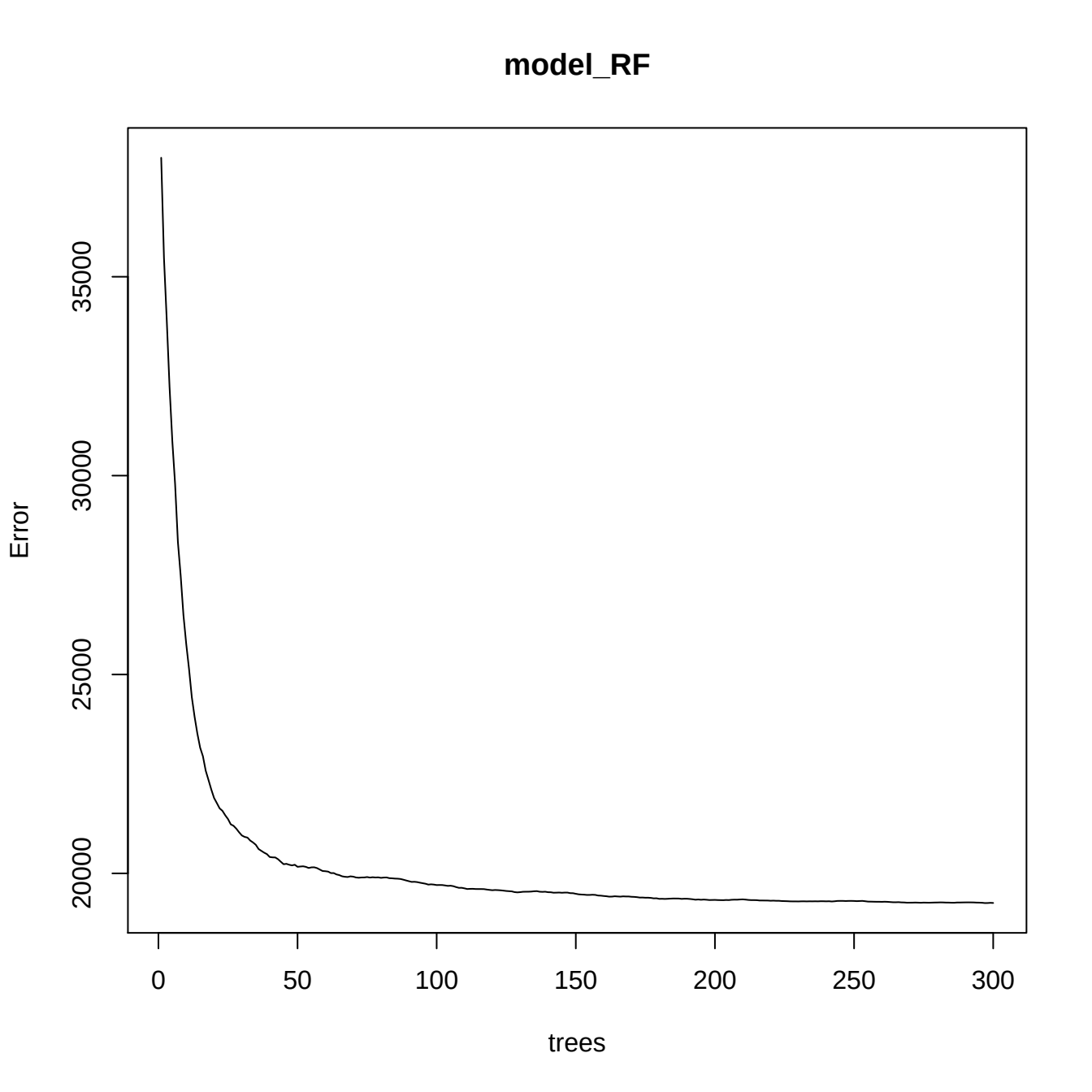

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
2
4
6
8
10
12
−4
−2
0
2
4


Histogram of df1$month
Frequency
0
500
1000
1500
2000
2
4
6
8
10
12
df1$month


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
2009
2010
2011
2012
2013
2014
2015
−4
−2
0
2
4


Histogram of df1$year
Frequency
0
500
1000
1500
2009
2010
2011
2012
2013
2014
2015
df1$year


282
n=9724 100%
yes
dropoff_latitude < 41
no
189
n=599 6%
pickup_latitude >= 41
288
n=9125 94%
pickup_latitude < 41
293
n=8760 90%
dropoff_longitude >= −74
251
n=1151 12%
pickup_longitude < −74
300
n=7609 78%
pickup_longitude >= −74
239
n=610 6%
dropoff_longitude < −74
86
n=275 3%
276
n=324 3%
161
n=365 4%
116
n=378 4%
317
n=773 8%
110
n=254 3%
331
n=356 4%
305
n=6999 72%


model_RF
Error
20000
25000
30000
35000
0
50
100
150
200
250
300
trees