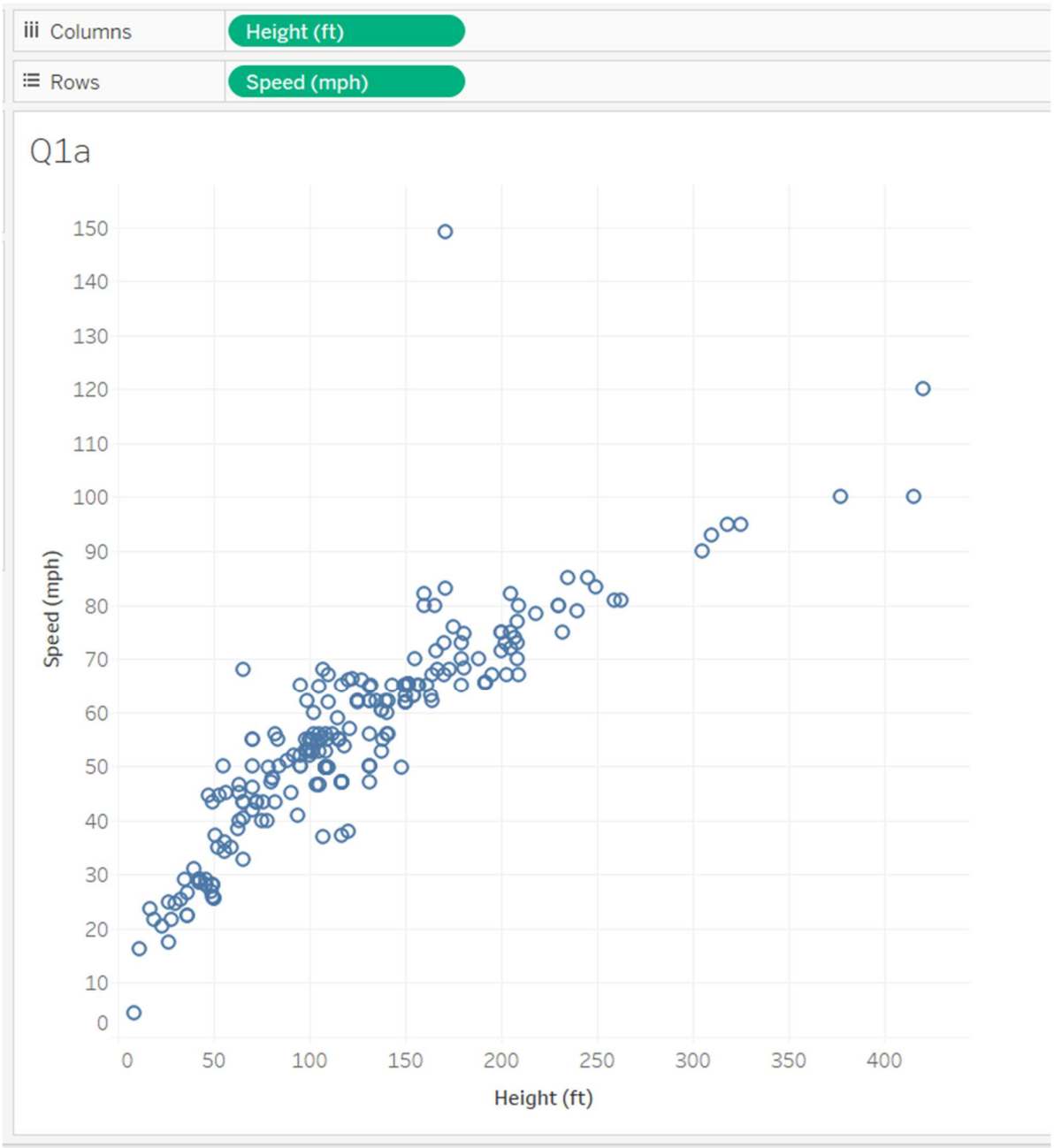
DATA ANALYTICS AND VISUALIZATION WITH TABLEAU

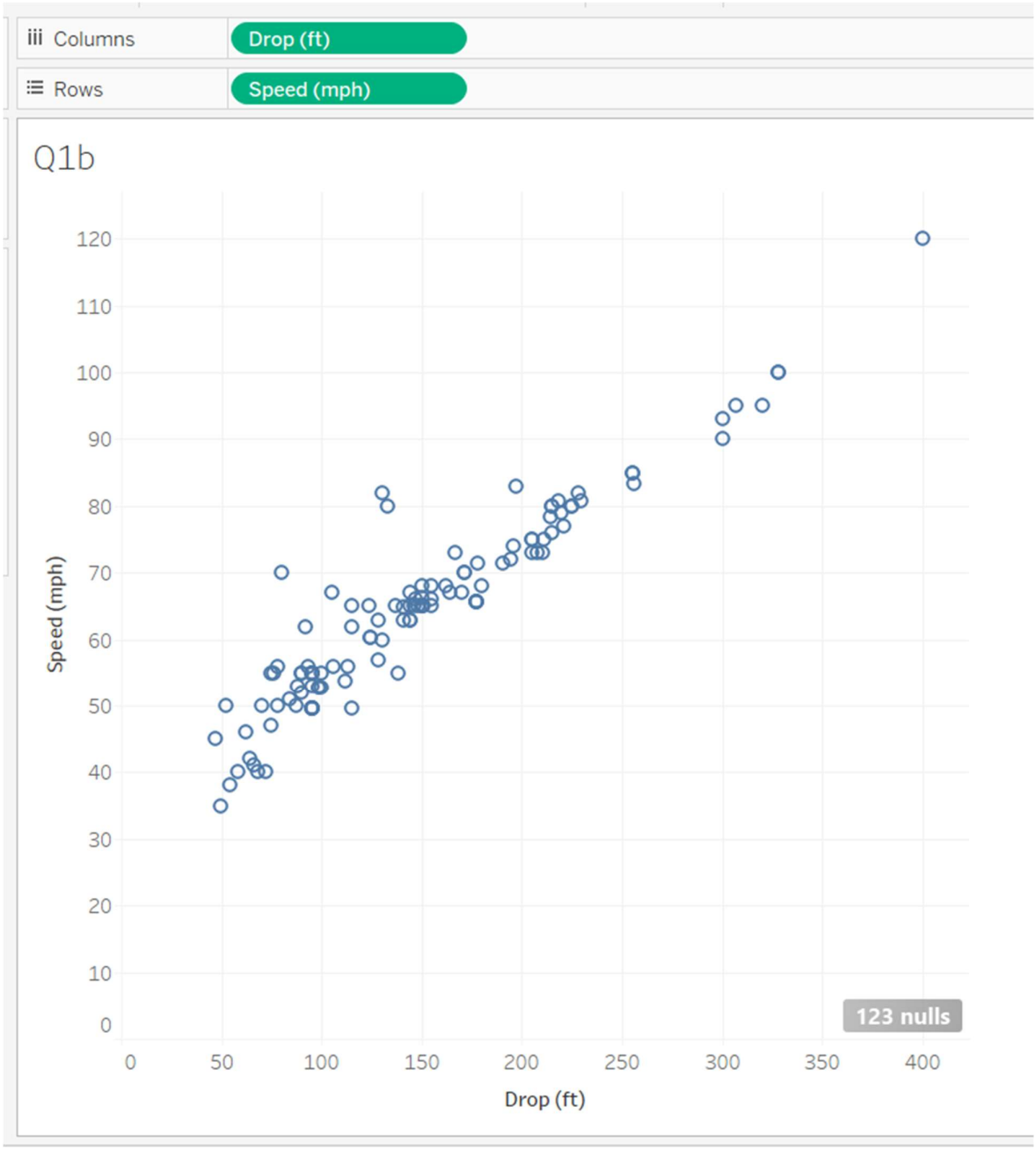
Assignment: Day 4

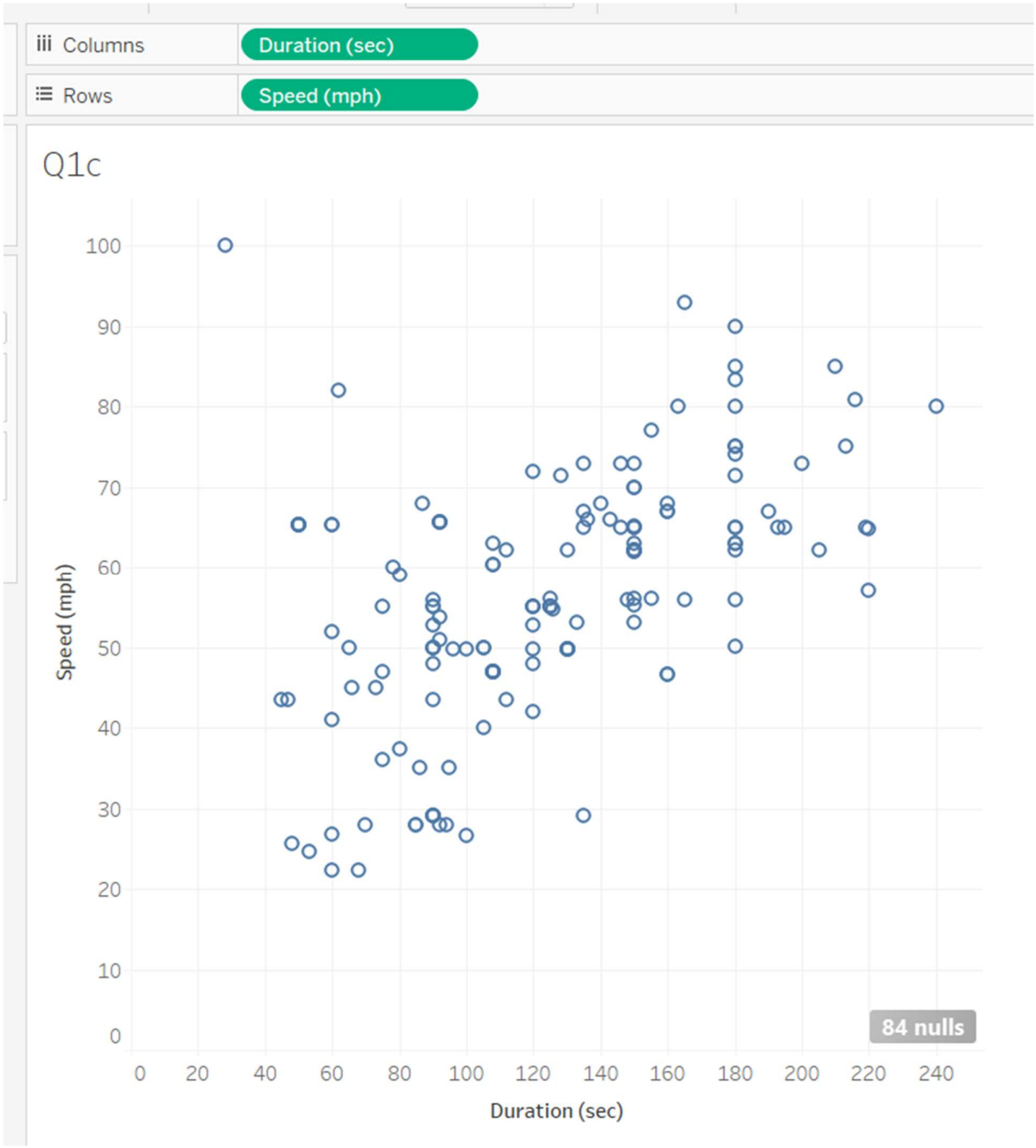
Rahul S Bhat

PES2UG19CS315

7th Semester

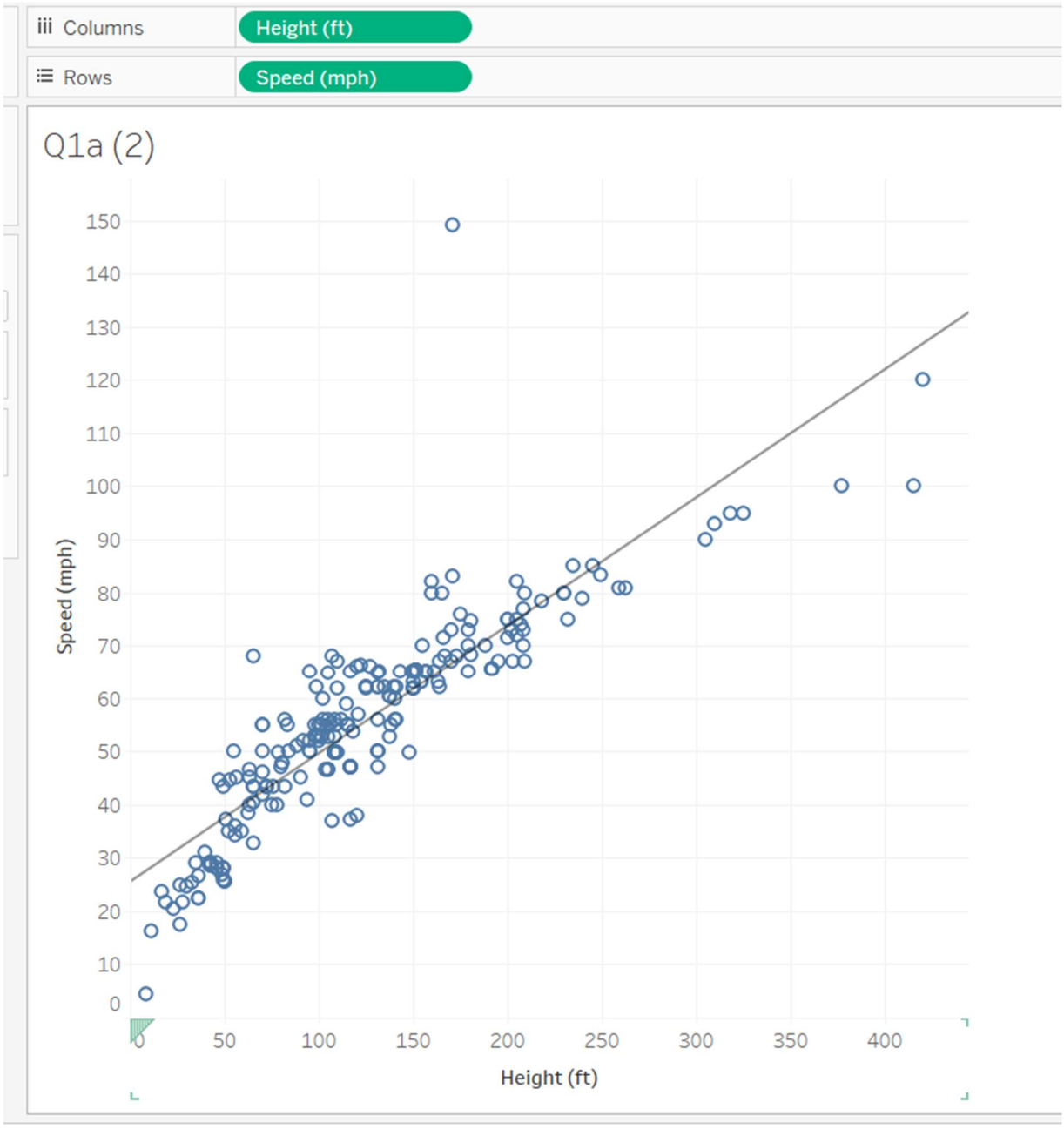


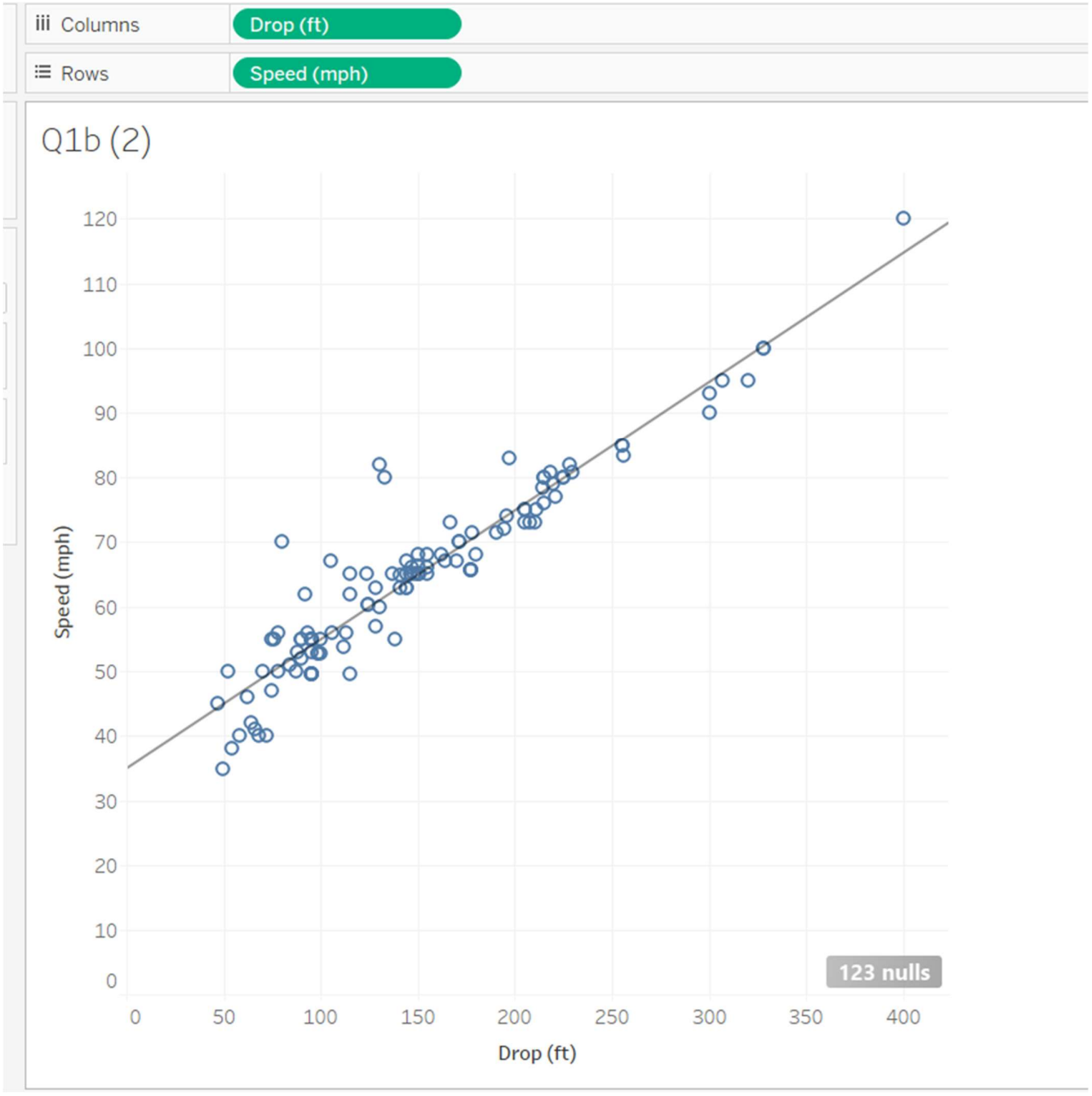

Columns
Drop (ft)
Rows
Speed (mph)
Q1b
120
110
100
90
80
70
60
50
40
30
20
10
0
Speed (mph)
0
50
100
150
200
250
300
350
400
Drop (ft)
123 nulls



Question One: Evaluate the regression conditions for each plot. Explain why or why not it is appropriate to run a regression analysis on each plot. Please address all conditions covered in this module.

It is appropriate to run regression analysis on plots Q1a and Q1b as the points are not too scattered and can approximately fit a straight line. Q1c cannot be used for regression analysis as the data is all scattered and doesn't seem to fit a line or a curve and hence it is not possible to predict the relation among the two plotted variables.


Columns
Height (ft)
Rows
Speed (mph)
Q1a (2)
Speed (mph)
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0
50
100
150
200
250
300
350
400
Height (ft)



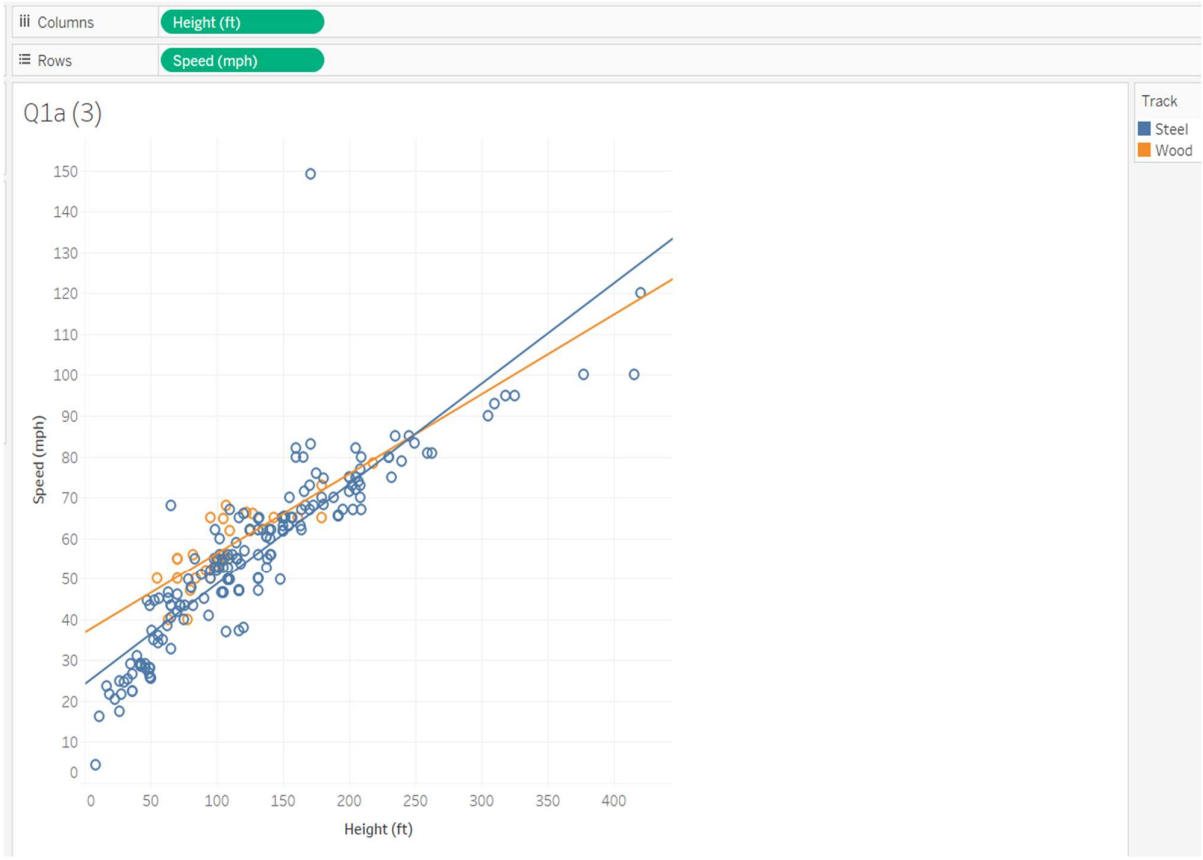
Question Two: Which variable(s) can you use to predict Speed?

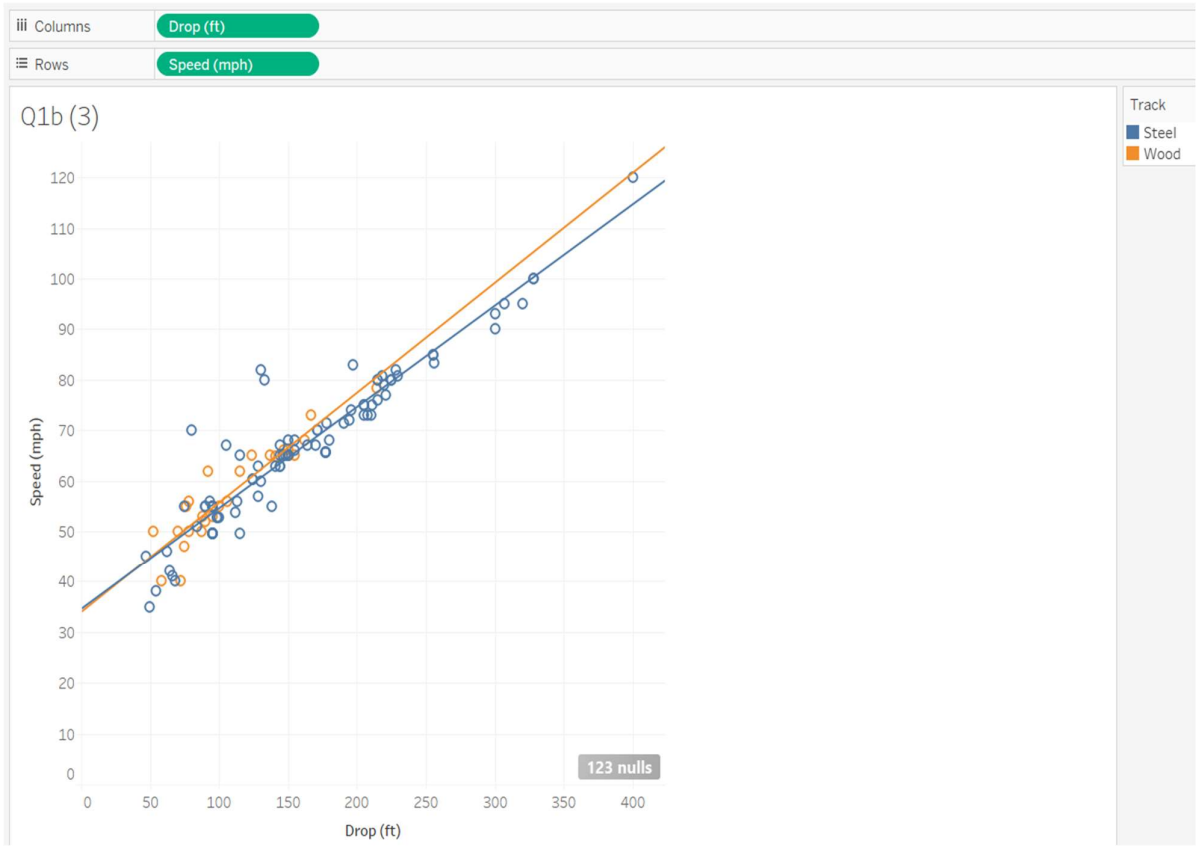
Drop can be used to predict speed as it fits the linear trend line better. (also higher R-Squared value)

Question Three: What is the regression equation for the trend line(s)?

For Q1a: Speed (mph) = 0.241085*Height (ft) + 25.5248

For Q1b: Speed (mph) = 0.199285*Drop (ft) + 34.9793


Columns
Height (ft)
Rows
Speed (mph)
Q1a (3)
Track
Steel
Wood
Speed (mph)
Height (ft)


Columns
Drop (ft)
Rows
Speed (mph)
Q1b (3)
Track
Steel
Wood
Speed (mph)
Drop (ft)
123 nulls

Question Four: Are there any differences between steel and wooden coasters in your analyses? Explain in few sentences.

There are differences.

Q1a(3)	Q1b(3)
Wood Speed (mph) = 0.195034*Height (ft) + 36.7227 R-Squared: 0.66751 P-value: < 0.0001 Steel Speed (mph) = 0.245764*Height (ft) + 24.0895 R-Squared: 0.785368 P-value: < 0.0001	Wood Speed (mph) = 0.217476*Drop (ft) + 33.9118 R-Squared: 0.848486 P-value: < 0.0001 Steel Speed (mph) = 0.200384*Drop (ft) + 34.5256 R-Squared: 0.890029 P-value: < 0.0001

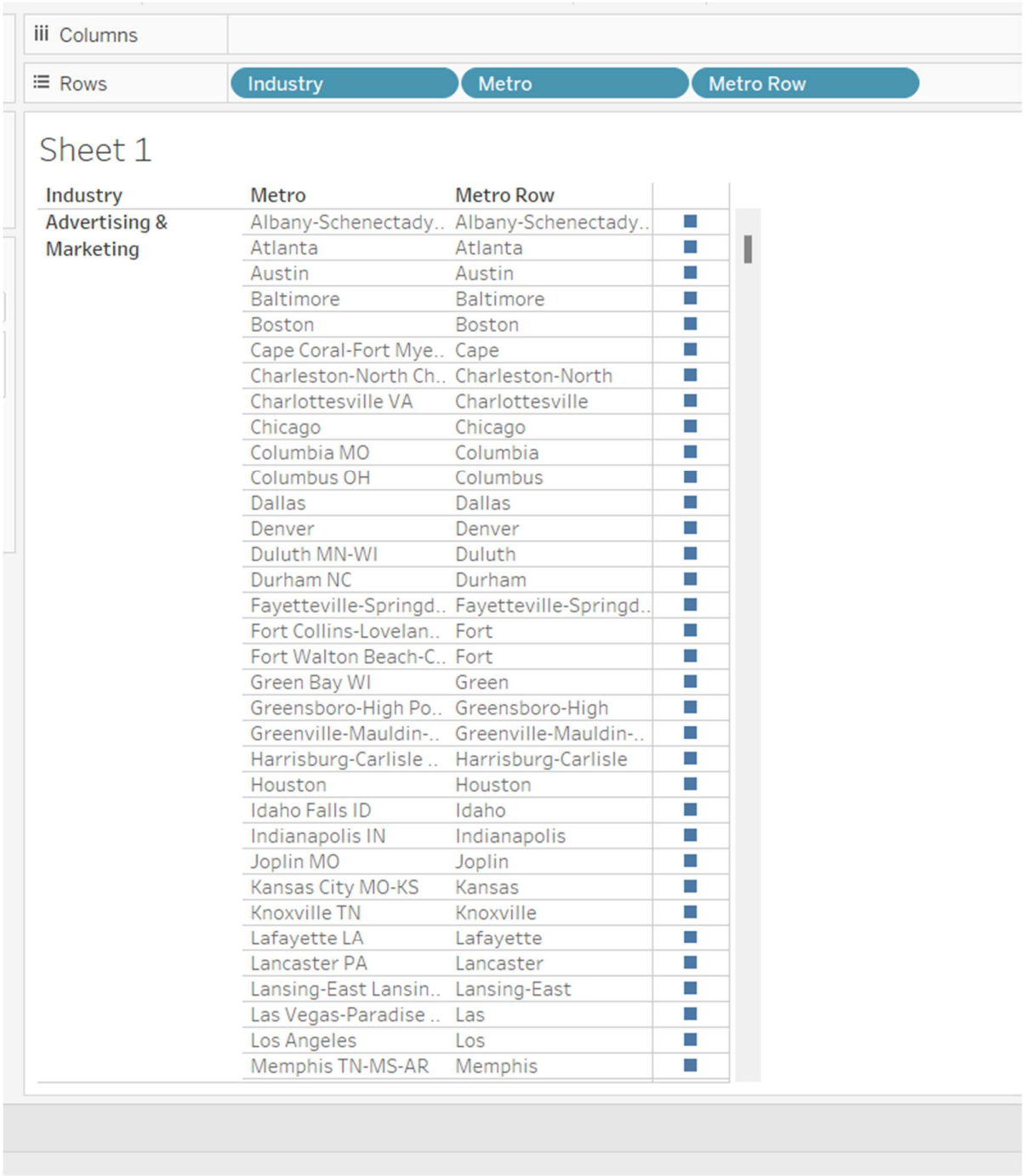
For Q1a(3), R-squared value for Steel is greater and hence we can say that the model fits better for steel category

Similarly in case of Q1b(3) R-squared value is greater for Steel and hence we can say that the model fits better for steel category and the more precisely, we can use drop values under steel track category to predict the speed

Question Five

Demonstrate the following

row calculation



Industry	Metro	Metro Row
Advertising & Marketing	Albany-Schenectady..	Albany-Schenectady..
	Atlanta	Atlanta
	Austin	Austin
	Baltimore	Baltimore
	Boston	Boston
	Cape Coral-Fort Mye..	Cape
	Charleston-North Ch..	Charleston-North
	Charlottesville VA	Charlottesville
	Chicago	Chicago
	Columbia MO	Columbia
	Columbus OH	Columbus
	Dallas	Dallas
	Denver	Denver
	Duluth MN-WI	Duluth
	Durham NC	Durham
	Fayetteville-Springd..	Fayetteville-Springd..
	Fort Collins-Lovelan..	Fort
	Fort Walton Beach-C..	Fort
	Green Bay WI	Green
	Greensboro-High Po..	Greensboro-High
	Greenville-Mauldin-..	Greenville-Mauldin-..
	Harrisburg-Carlisle ..	Harrisburg-Carlisle
	Houston	Houston
	Idaho Falls ID	Idaho
	Indianapolis IN	Indianapolis
	Joplin MO	Joplin
	Kansas City MO-KS	Kansas
	Knoxville TN	Knoxville
	Lafayette LA	Lafayette
	Lancaster PA	Lancaster
	Lansing-East Lansin..	Lansing-East
	Las Vegas-Paradise ..	Las
	Los Angeles	Los
	Memphis TN-MS-AR	Memphis

calculate field

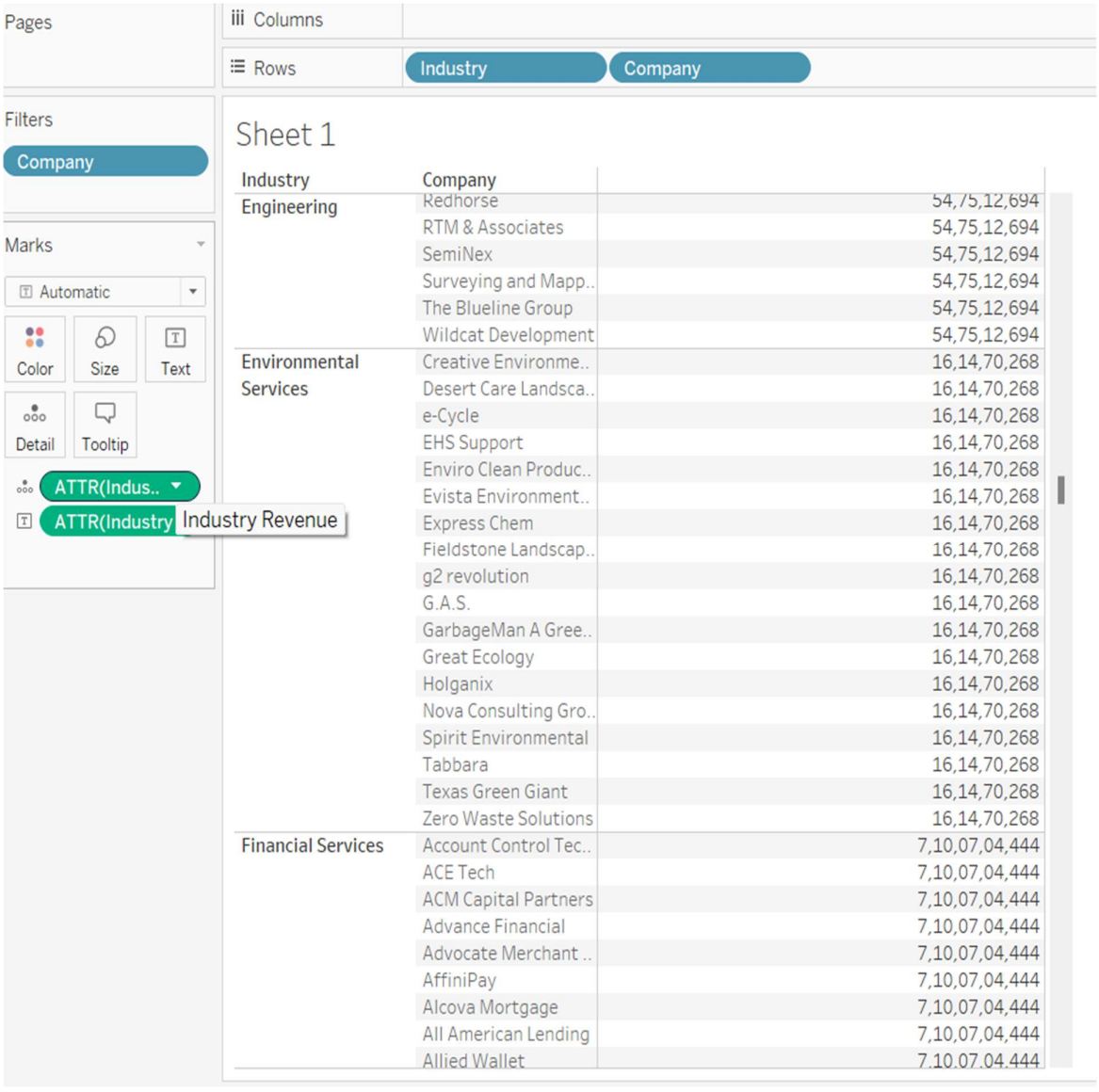
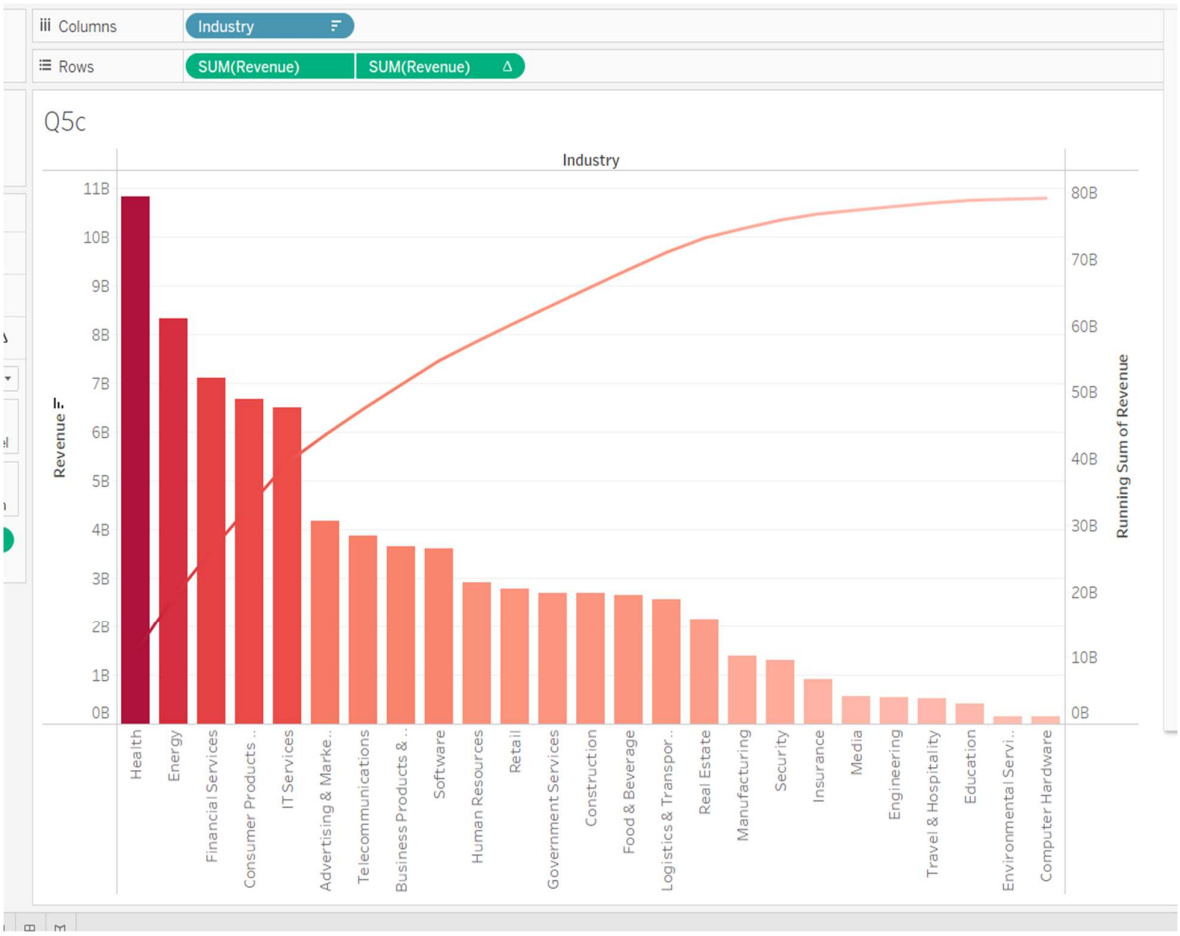
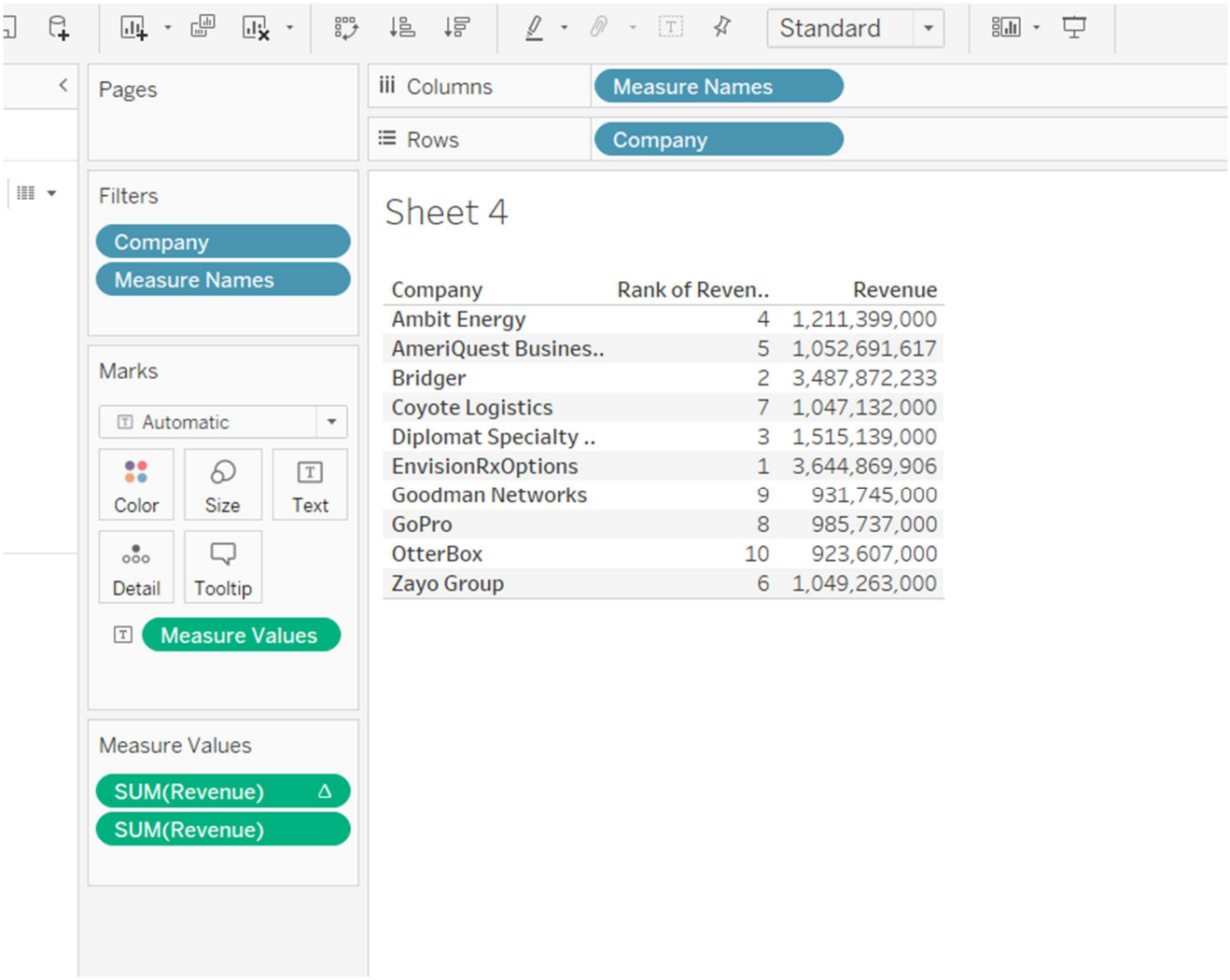


table calculation



ranking



Company	Rank of Reven..	Revenue
Ambit Energy	4	1,211,399,000
AmeriQuest Busines..	5	1,052,691,617
Bridger	2	3,487,872,233
Coyote Logistics	7	1,047,132,000
Diplomat Specialty ..	3	1,515,139,000
EnvisionRxOptions	1	3,644,869,906
Goodman Networks	9	931,745,000
GoPro	8	985,737,000
OtterBox	10	923,607,000
Zayo Group	6	1,049,263,000