

# Error Analysis on Select Clause

Zihua

# Column Level Performance

**Without similarity between column header and question**

	Pred = 1	Pred = 0
Label = 1	236	37
Label = 0	578	1195

	Pred = 1	Pred = 0
Label = 1	50	16
Label = 0	116	377

**With similarity between column header and question**

	Pred = 1	Pred = 0
Label = 1	242	31
Label = 0	425	1348

	Pred = 1	Pred = 0
Label = 1	54	12
Label = 0	107	386

# Table Level Performance

Total

238

Exactly match

38

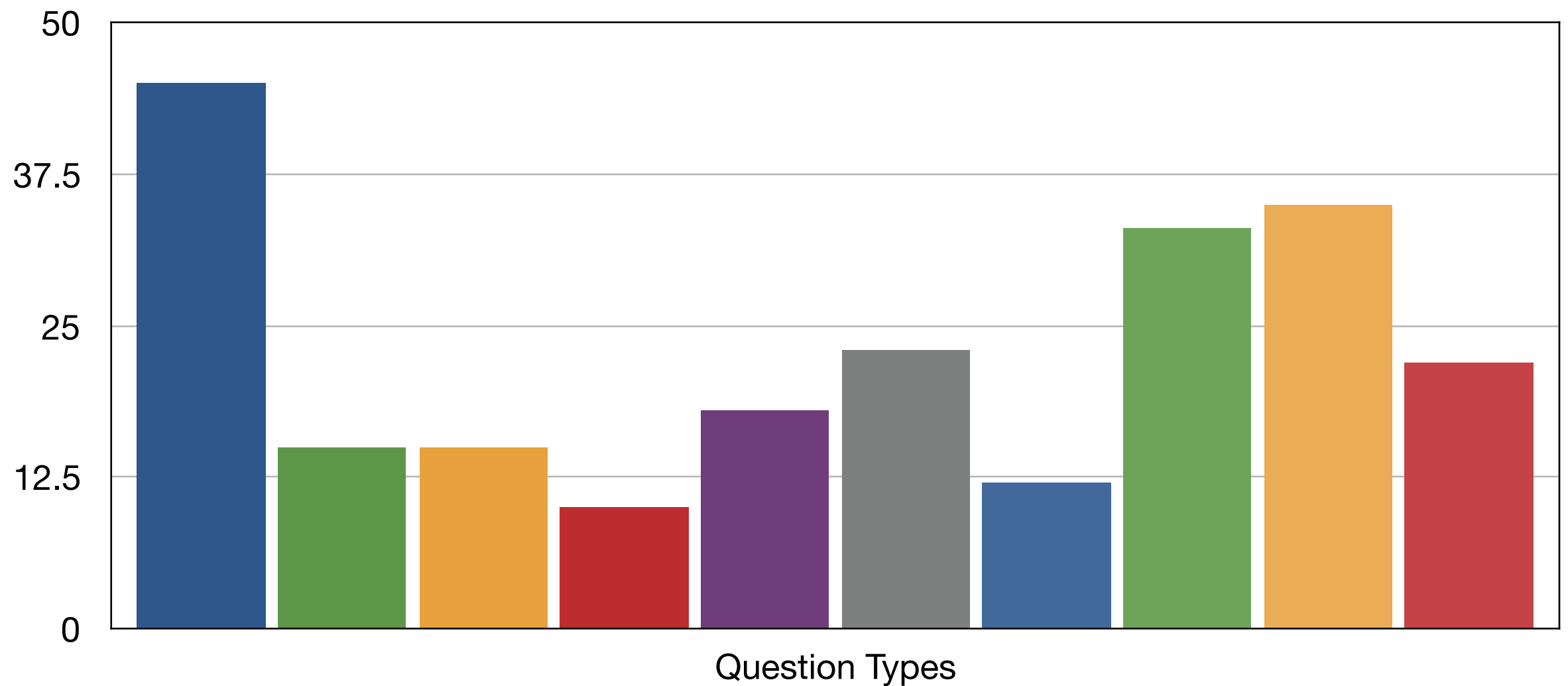
Ground Truth is a subset of prediction

170

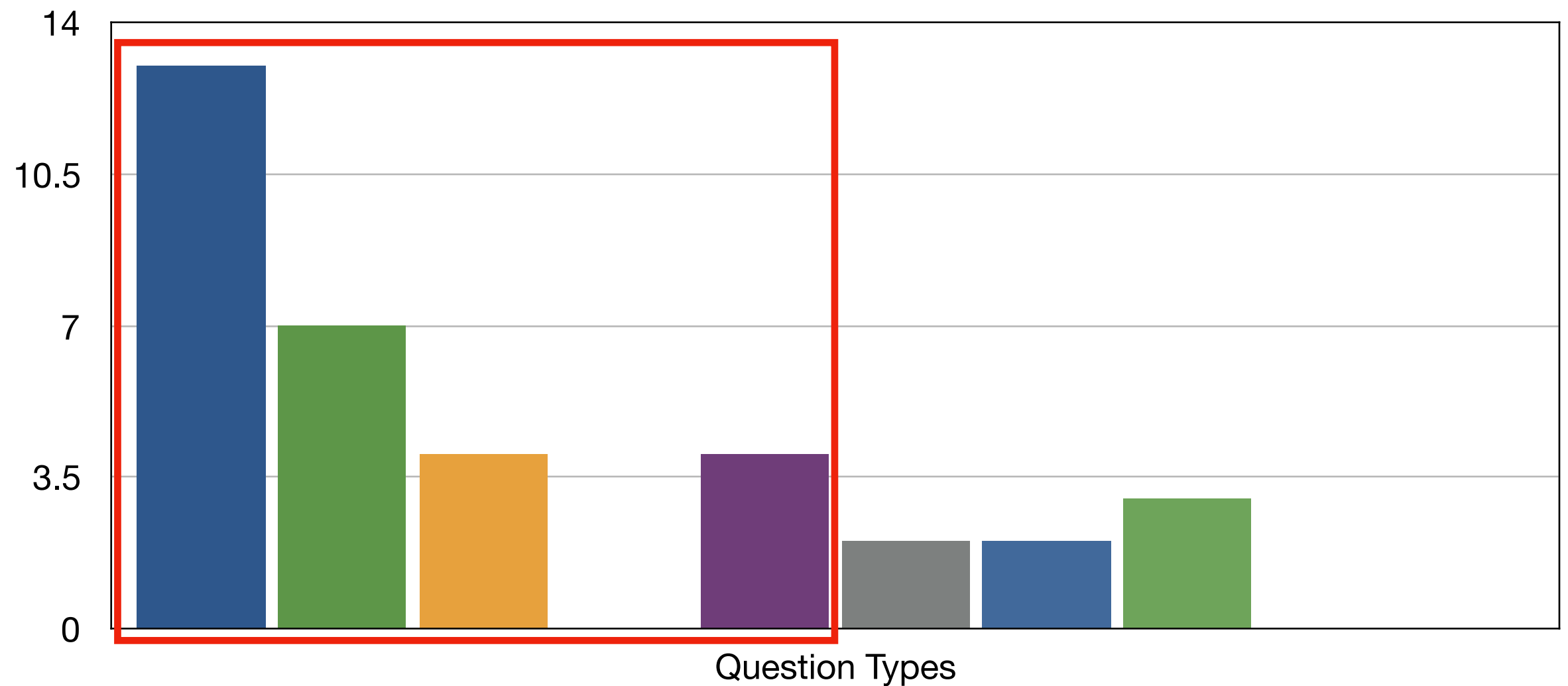
Missing some select column

30

# Error Analysis

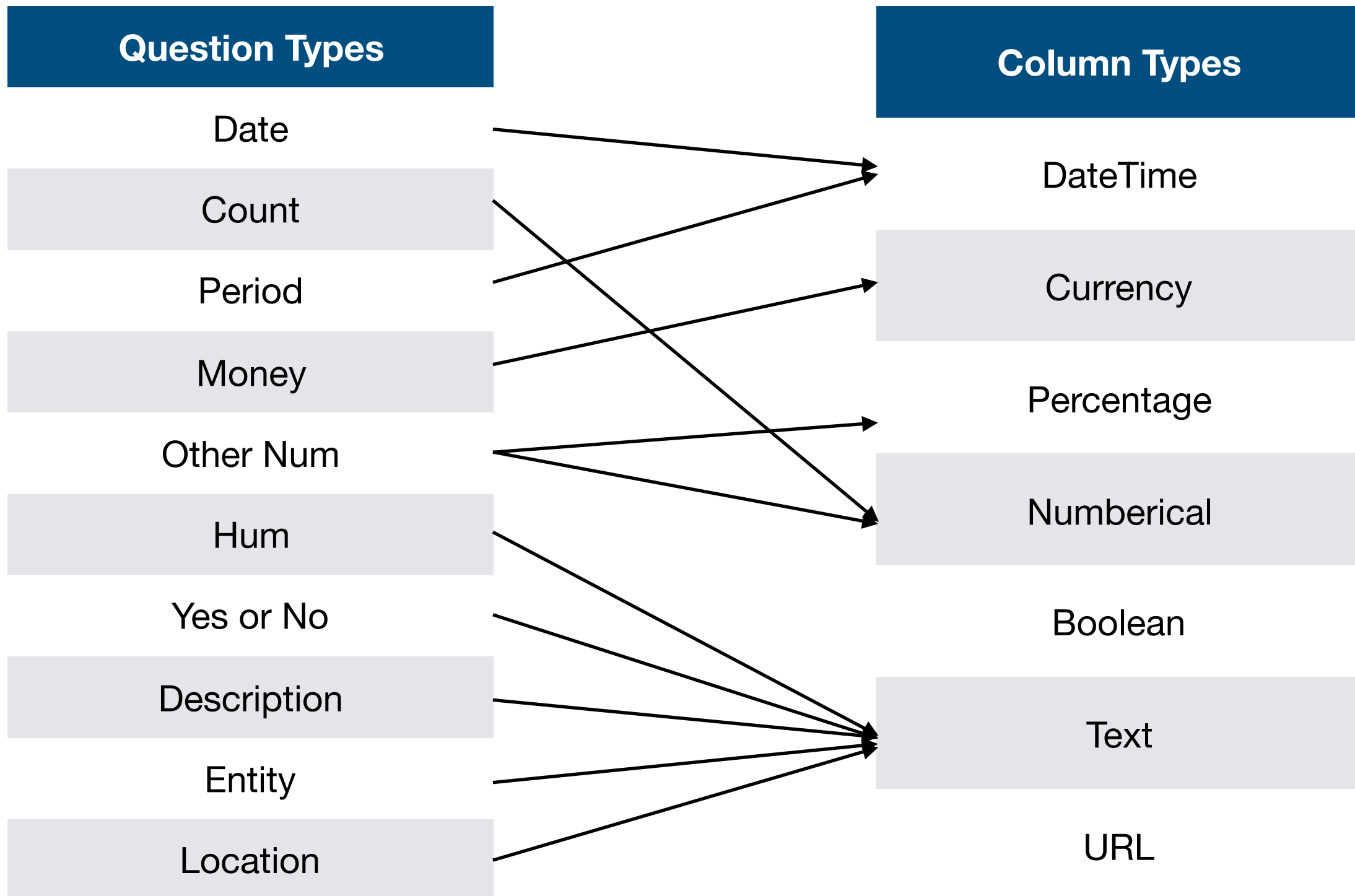


# Error Analysis



**Have corresponding column types**

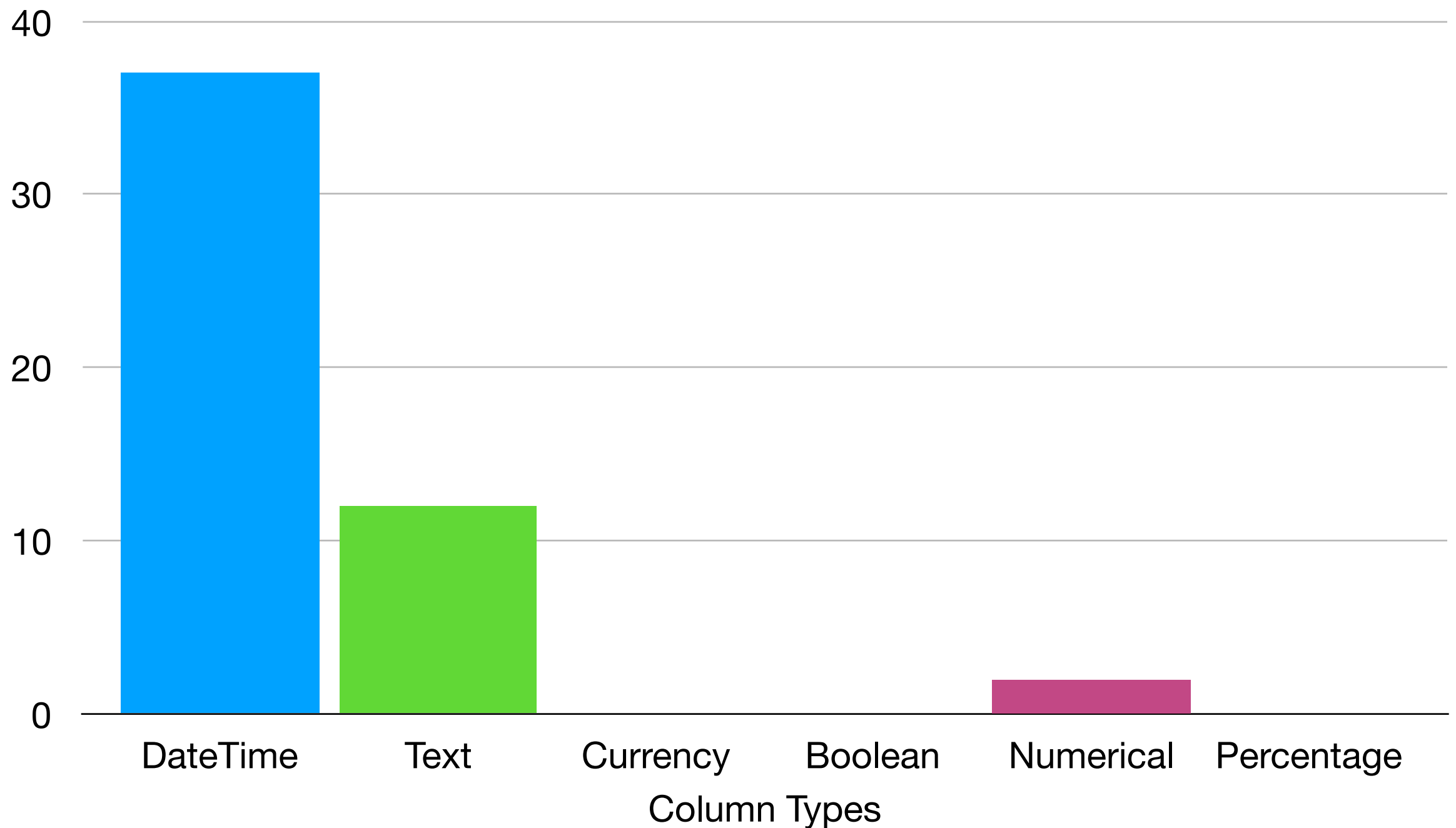
# Error Analysis



**Error on column  
types recognition**

# Question Type: date

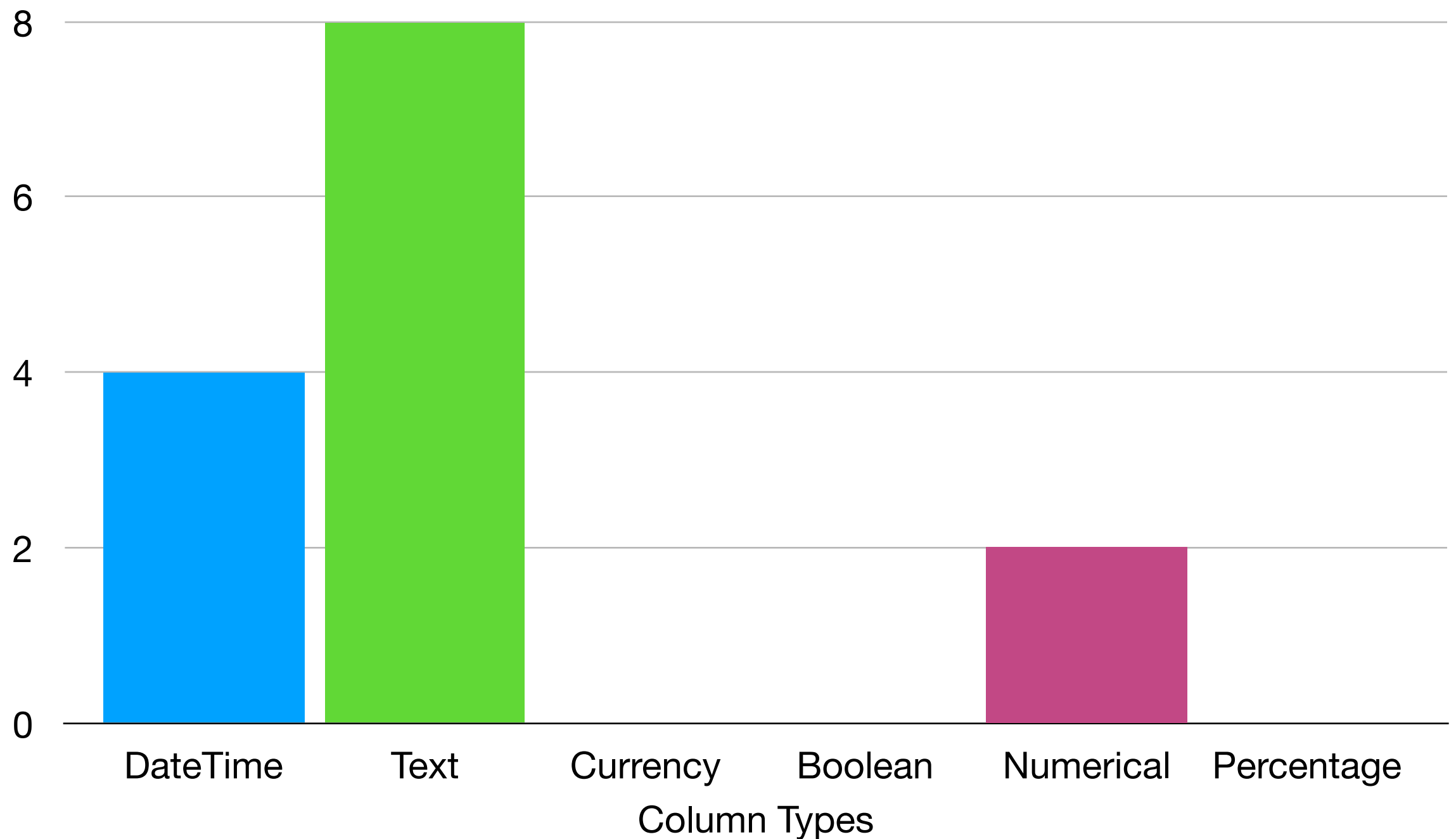
DateTime Text Currency Boolean Numerical  
Percentage





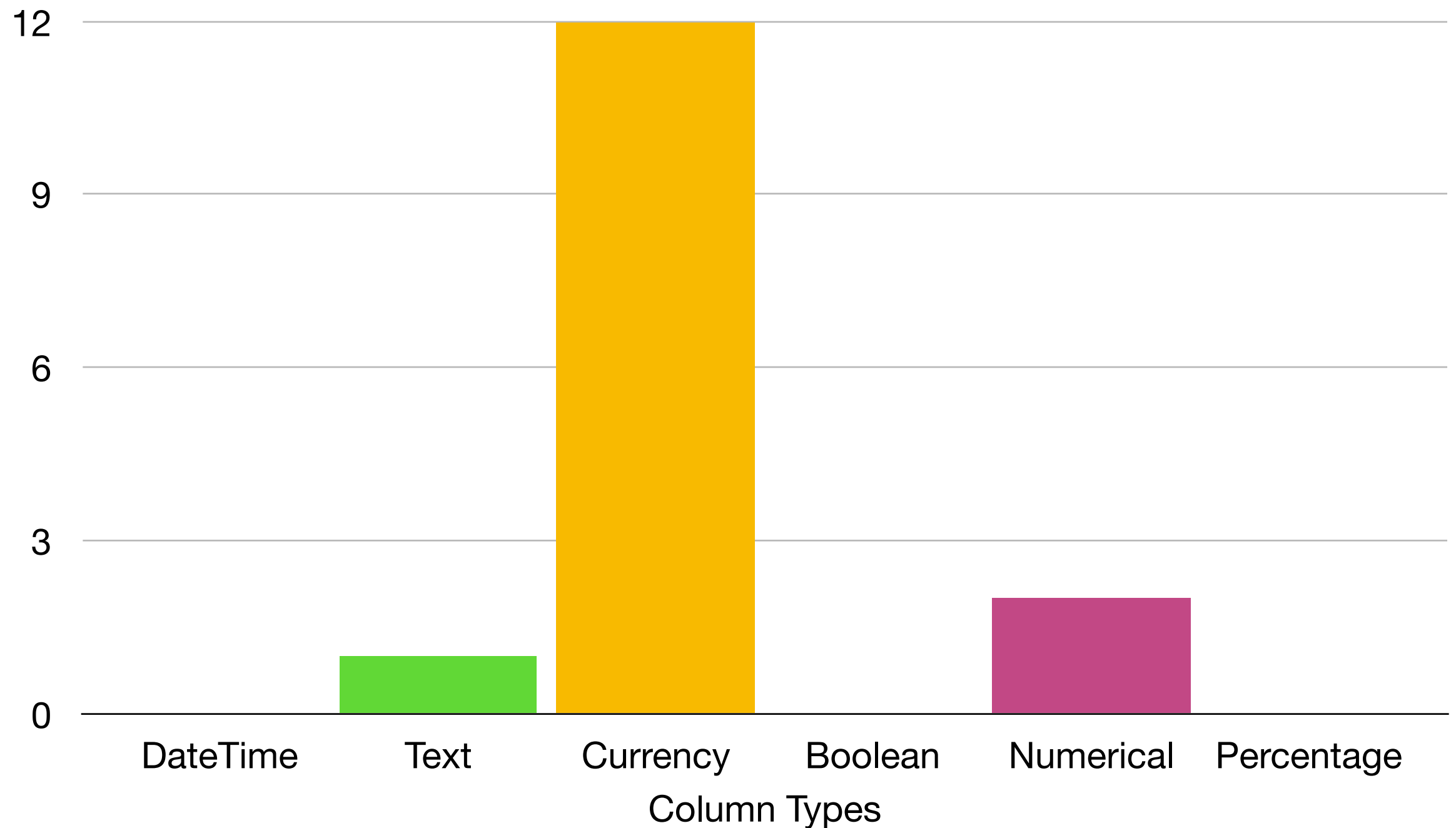
# Question Type: period

DateTime Text Currency Boolean Numerical  
Percentage



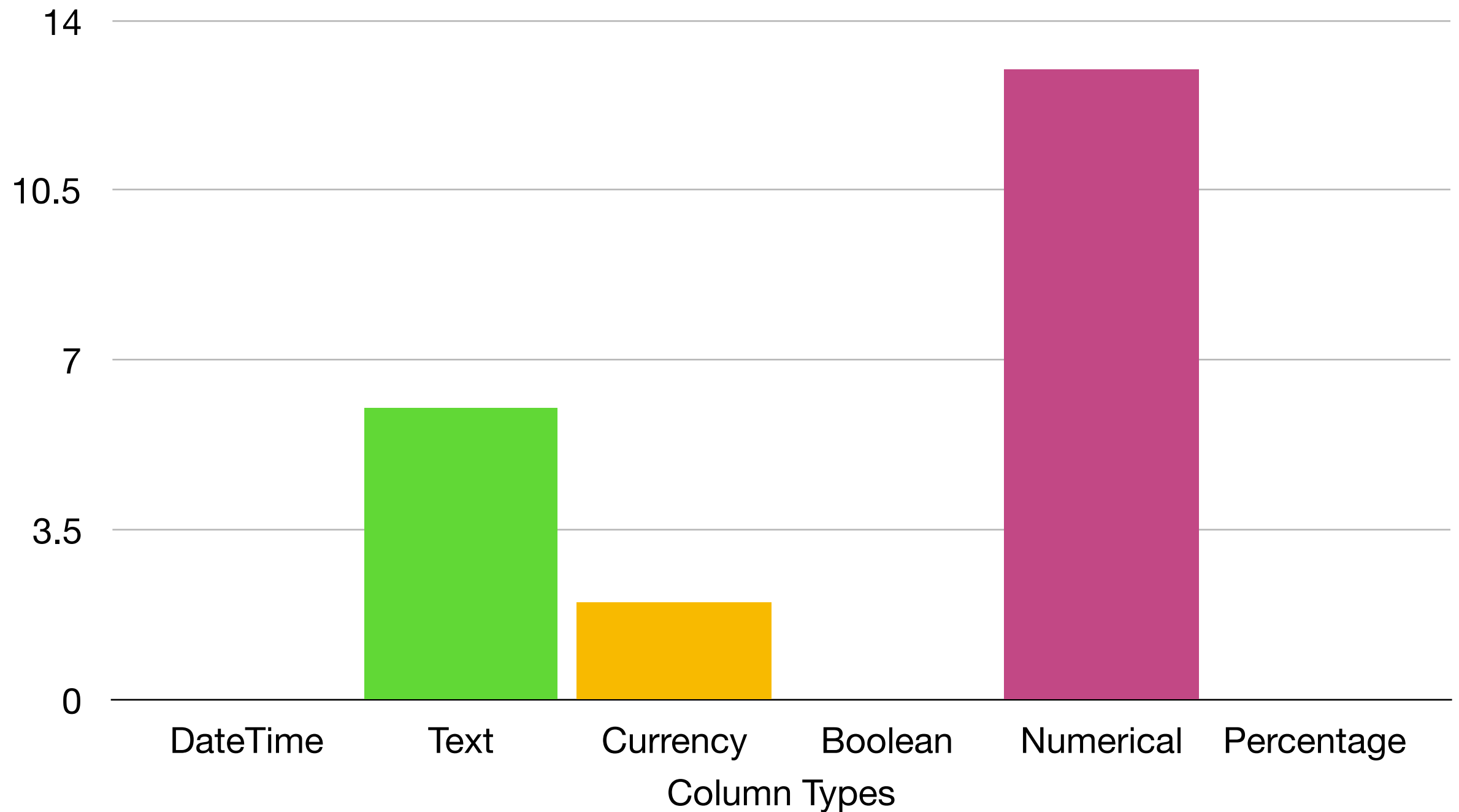
# Question Type: money

■ DateTime   ■ Text   ■ Currency   ■ Boolean   ■ Numerical  
■ Percentage



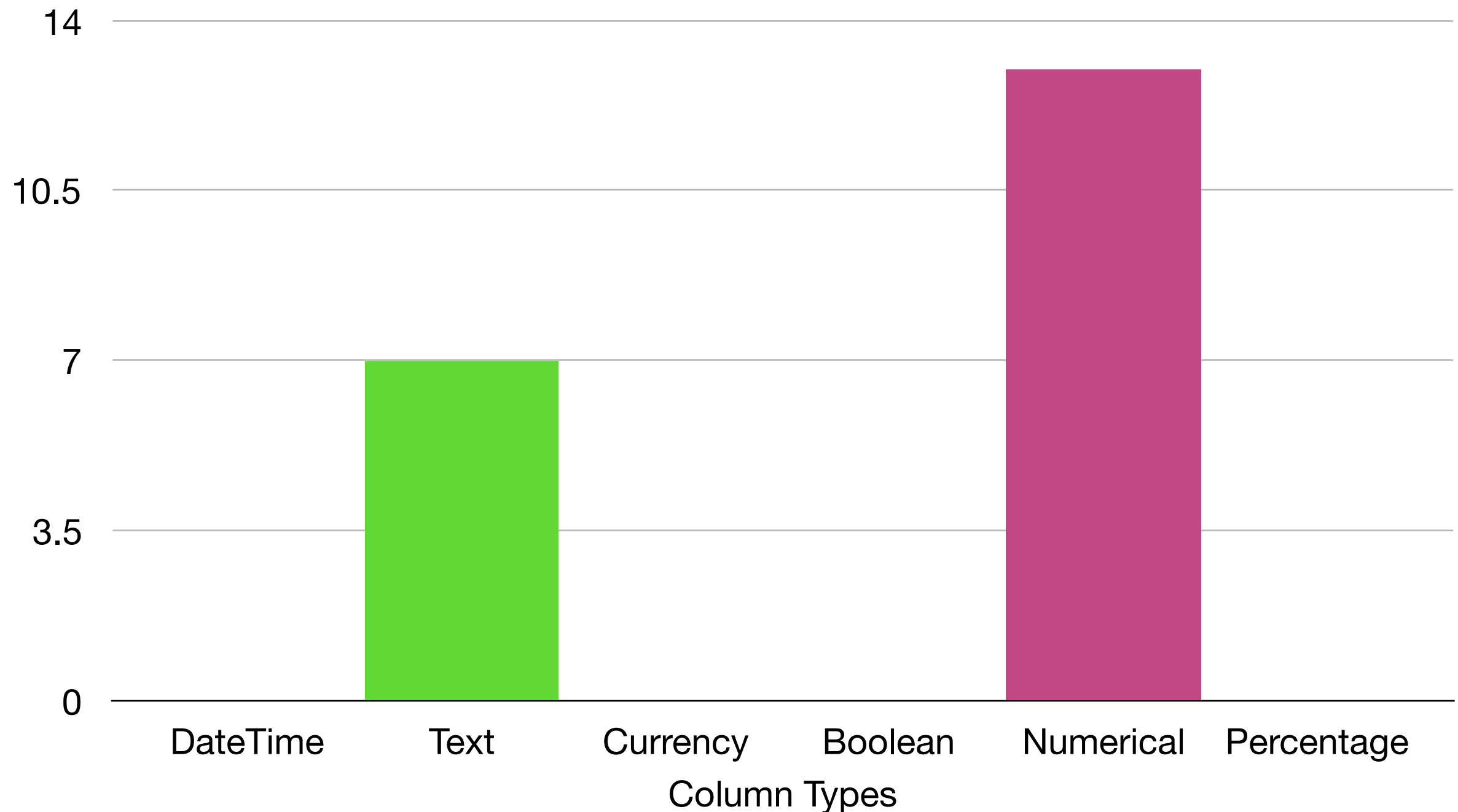
# Question Type: count

DateTime Text Currency Boolean Numerical  
Percentage



# Question Type: other NUM

DateTime Text Currency Boolean Numerical  
Percentage



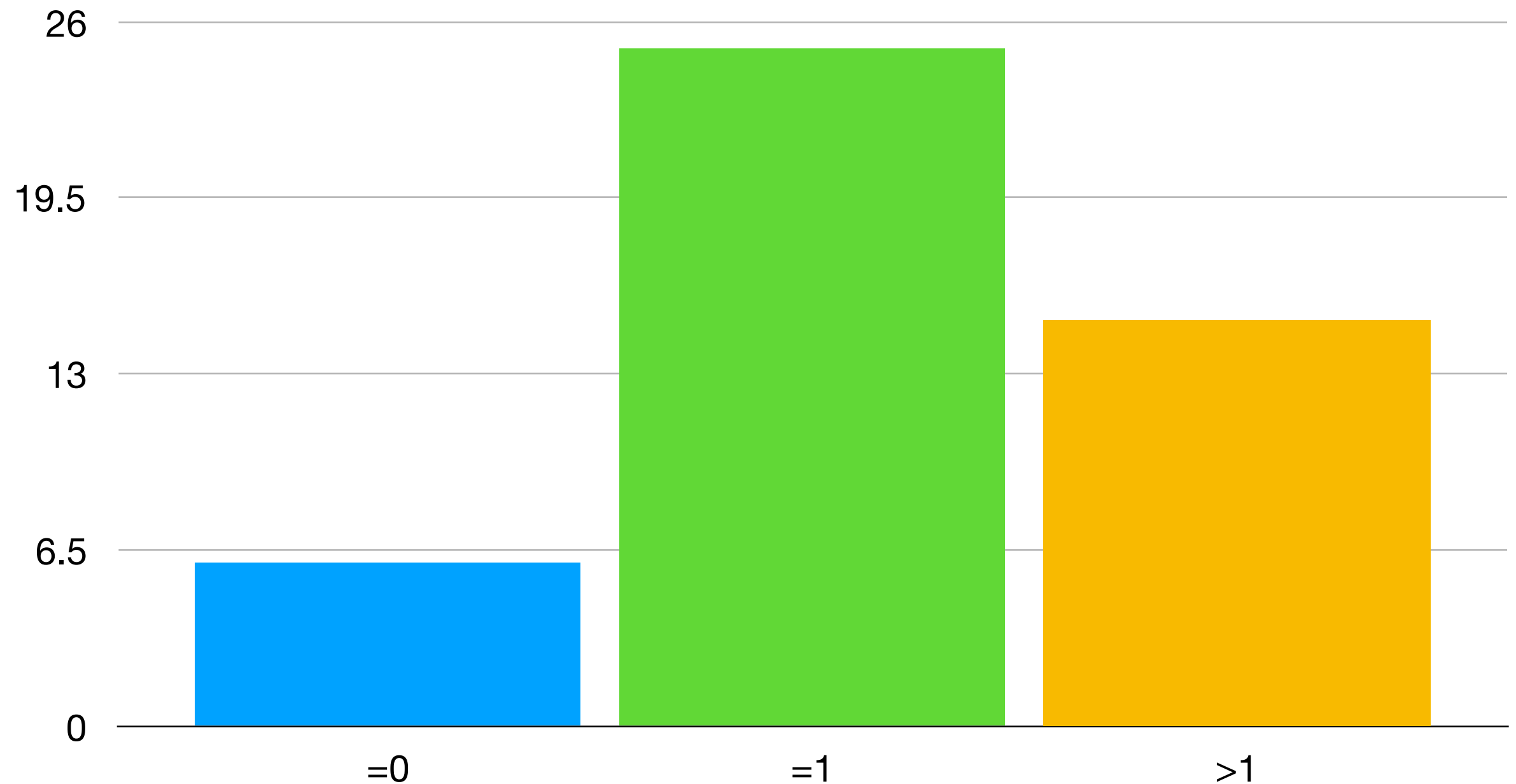
**Error on multiple  
columns with same type**

# Question Type: date

■ =0

■ =1

■ >1

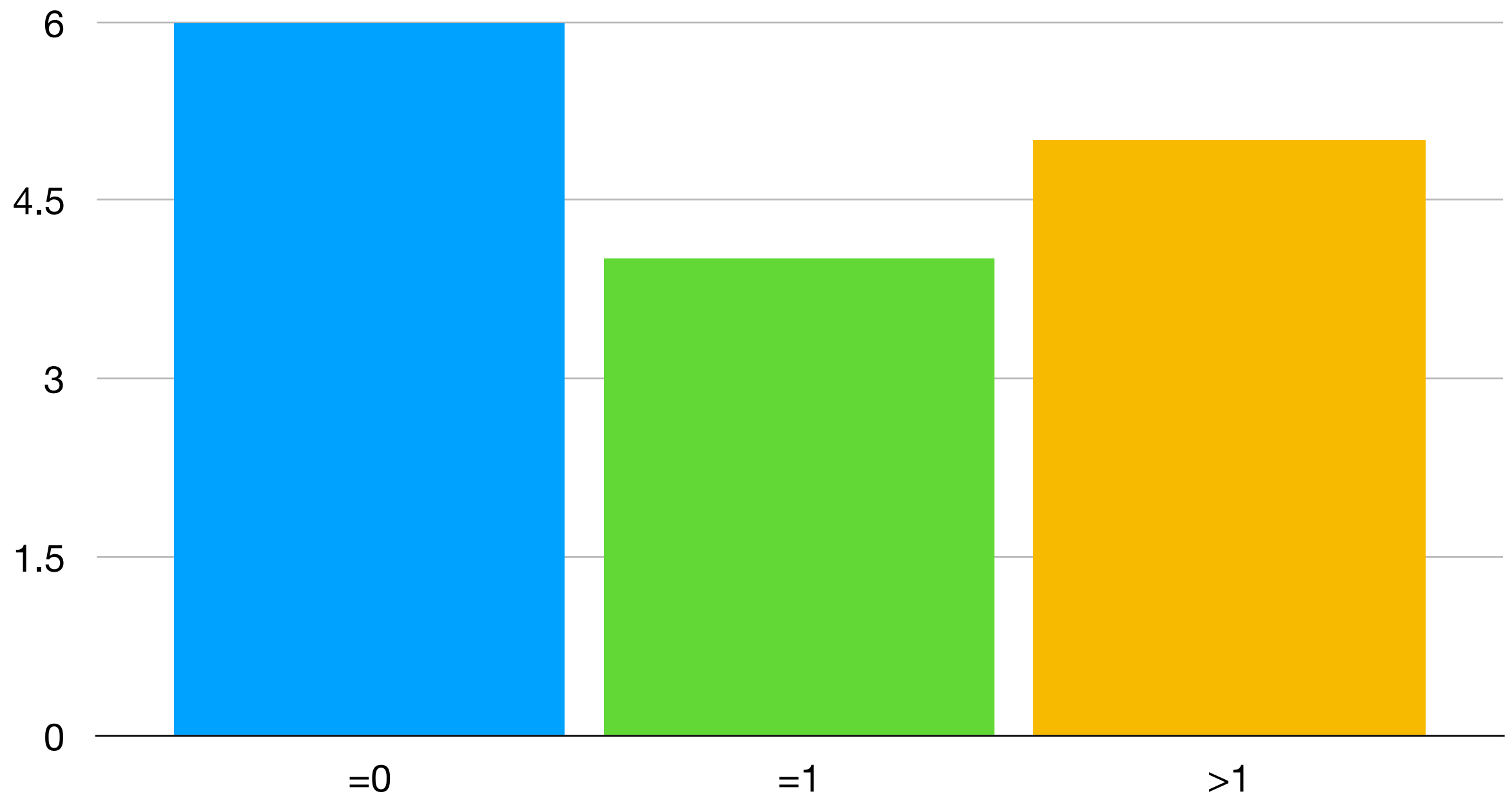


# Question Type: period

■ =0

■ =1

■ >1

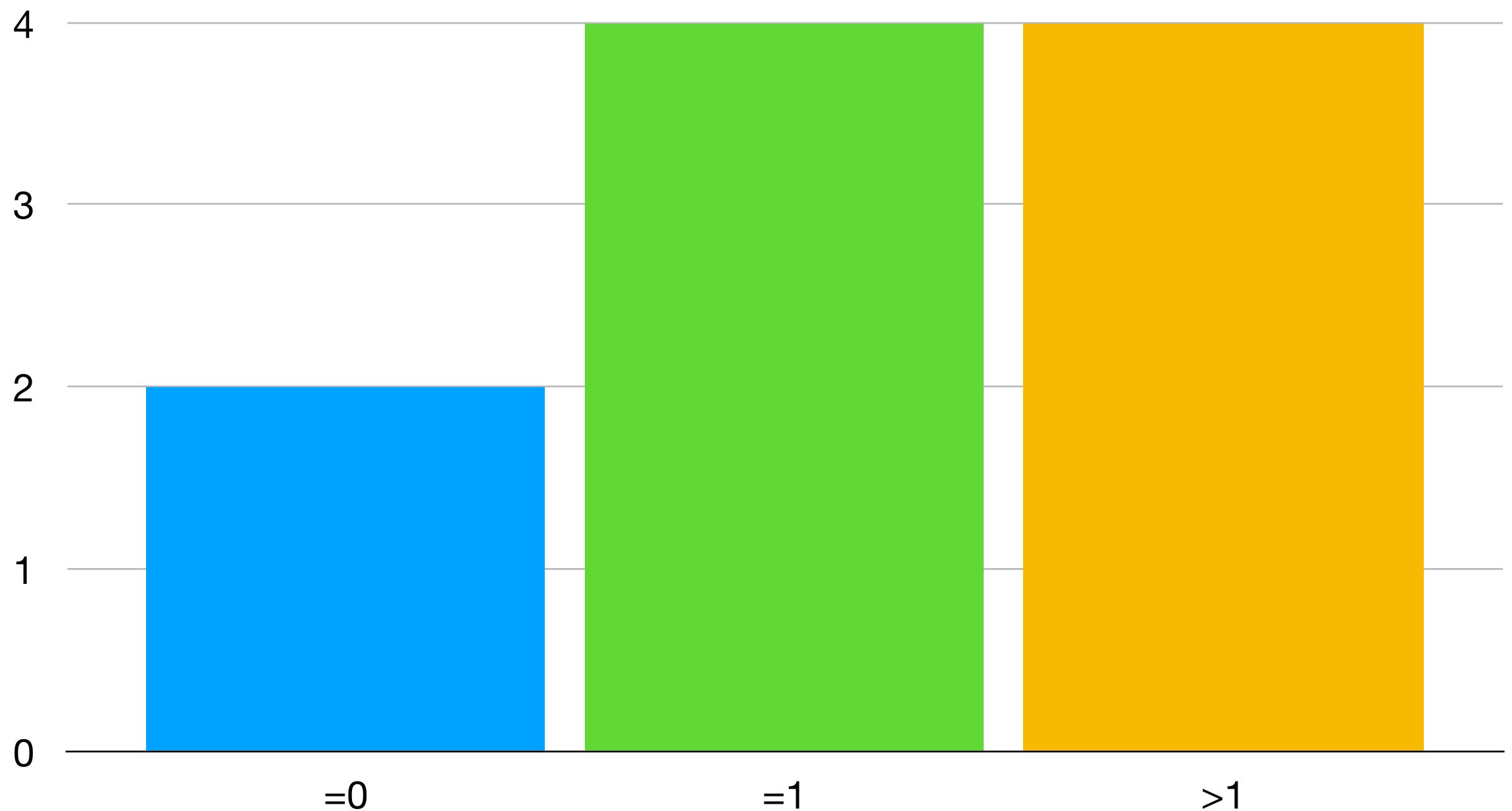


# Question Type: money

■ =0

■ =1

■ >1



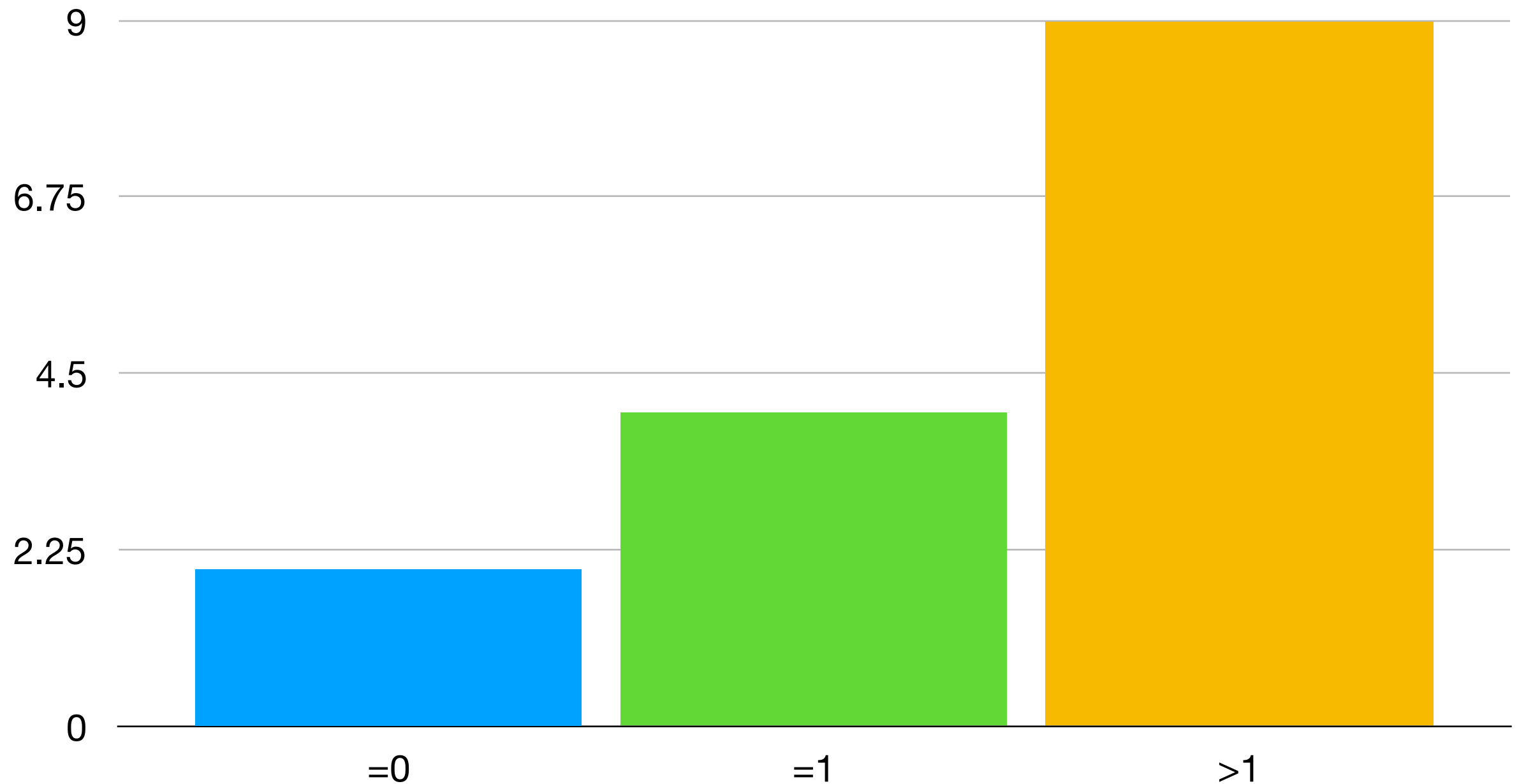


# Question Type: count

■ =0

■ =1

■ >1



# Question Type: other NUM

■ =0

■ =1

■ >1

